

SPREADSHEET ACTIVITIES FOR A DATA MANAGEMENT  
UNIT: A TEACHERS' RESOURCE USING  
Microsoft Excel

CENTRE FOR NEWFOUNDLAND STUDIES

**TOTAL OF 10 PAGES ONLY  
MAY BE XEROXED**

(Without Author's Permission)

RALPH PERCEY







Spreadsheet Activities for a Data Management Unit:

A Teachers' Resource using *Microsoft Excel*

A project submitted to the Faculty of Education

in partial fulfillment of the requirements

for the degree of

Masters of Education, Teaching and Learning

Ralph Percey

Memorial University of Newfoundland

2000

## Acknowledgments

I would like to begin by thanking Dr. John Grant McLoughlin for his role as supervisor throughout my project. From assisting with the topic selection to directing and editing the final copy, Dr. Grant McLoughlin has been there. Feedback was always prompt, informative, and open for discussion. I always felt in control of the project with Dr. Grant McLoughlin there to provide guiding words of wisdom. His enthusiasm, readiness to meet on short notice and overall approach to supervision made the final completion of the project a welcome reality. I again offer my many thanks to Dr. Grant McLoughlin.

I would also like to thank my loving wife Susan. Thank you for your love, support, help, patience, and understanding during both my graduate and undergraduate studies. In more ways than you will ever know, you have been the inspiration that led me to these degrees. Thank you for our two beautiful sons Devin and Dylan and the sacrifices you have made and are yet to make. I assure you that nothing you have sacrificed has gone unnoticed or unappreciated. I dedicate this project to you, Devin, and Dylan as a symbol of my thanks and appreciation.

I would like to thank and pay tribute to my parents James and Catherine Percey. The lessons you tried so hard to teach me did not all fall on deaf ears. Thanks for everything (except the procrastination gene Dad).

In closing, I offer a note of thanks to all those individuals (in-laws included) that have helped me to where I am today. No small piece of advice or words of wisdom are ever recklessly dismissed.

## Abstract

Statistics are all around us. Statistical reports from areas like business, weather, politics, and sports are encountered daily by students and adults. Yet, the study of statistics as a branch of mathematics has primarily been an optional course offering in the high school curriculum of Newfoundland. However, through the efforts of the Atlantic Provinces Education Foundation (APEF), the study of statistics will now have a compulsory place in each level of the regional high school mathematics curriculum. This curriculum also calls for the integration of appropriate technologies like computer software to explore the relevant objectives.

The aim of this project is to offer teachers a resource that supports the instruction of statistics through the integration of technology. This project is designed to explore various curriculum topics in statistics using the computer software package *Microsoft Excel*. The high school curriculum objectives for the Data Management unit as outlined in Course 1, Level Y, *Teacher's Pilot Draft Guidelines of the Atlantic Canada Mathematics Curriculum* (1997) provides the framework for the activities.

The project consists of a series of activities that use *Microsoft Excel* as a tool to help students investigate several concepts in statistics. The National Council of Teachers of Mathematics has long supported involving students in constructing and applying mathematical ideas, using a variety of instructional formats like small cooperative groups and individual explorations, and using computers as tools for learning and doing mathematics (NCTM, 1989, p. vi). This project integrates these ideas into an organized collection of activities.

## Table of Contents

Acknowledgments .....	ii
Abstract .....	iii
Introduction .....	1
Literature Review .....	3
The Need for Technology Integration .....	3
Why Use a Spreadsheet to Study Statistics? .....	5
Cooperative Learning and Reflective Practice .....	8
Overview of Project .....	12
Description .....	12
Classroom Climate .....	13
Note to the Teacher .....	14
Sequencing the Activities .....	15
Evaluation Plan .....	17
Data Management: A Set of Activities in <i>Excel</i> .....	19
Overview of Activity 1 .....	20
Activity 1 .....	21
Overview of Activity 2 .....	27
Activity 2 .....	28
Overview of Activity 3 .....	32
Activity 3 .....	33
Overview of Activity 4 .....	38
Activity 4 .....	39

Overview of Activity 5 .....	43
Activity 5 .....	44
Overview of Activity 6 .....	49
Activity 6 .....	50
References .....	54
Appendix A: Background to <i>Excel</i> .....	57
Background to <i>Excel</i> .....	58
Appendix B: Student Activity Sheets .....	61
Student Activity 1 .....	62
Student Activity 2 .....	65
Student Activity 3 .....	67
Student Activity 4 .....	70
Student Activity 5 .....	72
Student Activity 6 .....	75

## Introduction

The Atlantic Provinces Education Foundation (APEF) is an organization sponsored and managed by the governments of the four Atlantic Provinces. In addition to other subject areas, the APEF has been commissioned with the task of developing a common curriculum for mathematics. The pilot draft guidelines of the Atlantic Canada Mathematics Curriculum (1997, p. i) state that the National Council of Teachers of Mathematics (NCTM) *Curriculum and Evaluation Standards for School Mathematics* (1989) has greatly influenced this curriculum. This common curriculum now includes a single unit devoted to the study of statistics in each of Levels 1, 2, and 3 (grades 10, 11, and 12). The increased emphasis on statistics is a new focus for the Newfoundland high school mathematics curriculum that has previously included only an optional course in statistics at the high school level. This is inconsistent with all the other grade levels where statistics remains a core part of the curriculum.

I currently teach mathematics in an urban senior high school setting (Levels 1 to 3). After nine years of teaching, I have come to the conclusion that the study of statistics has not been given adequate attention in high school. My present school offers only one section of the optional Statistics 3104 course a year. This allows only 35 of 725 students to do this course. My discussions with mathematics teachers from other local urban schools confirm similar course offerings. I suspect that smaller schools would be even less likely to offer such an elective course. As statistics becomes a core component of the curriculum, teachers will be teaching topics with which they may have minimal recent experience. Furthermore, they likely have few resources available to support them in the implementation of the new curriculum.

The new draft *Principles and Standards Document* (NCTM, 1998) recommends that teachers have a background in mathematics well beyond that which they are required to teach. However, Hoffer (1982, p. 263) and Hogg (1992, p. 6) claim that many teachers do not have an adequate background in statistics and thus lack the knowledge to teach statistical topics. They note that this leads to a lack of ability and experience in selecting and integrating meaningful activities. No empirical evidence is offered to support these claims, however evidence of deficiencies in teachers' backgrounds in statistics can be found amongst present teachers. In conversations with several mathematics teaching colleagues, most reported that they have only completed the one or two statistics courses required by the degree program. This raises some concern over the issue of teacher preparedness to deliver the statistics topics in the APEF curriculum.

The APEF mathematics curriculum also calls for the integration of technology in meeting many of the objectives of the courses. In particular, I refer to the unit entitled Data Management from Course I, Level Y of the *Teacher's Pilot Draft Guidelines: Atlantic Canada Mathematics Curriculum*. The Key Stage Curriculum Outcomes 2, 3, and 4 call for the integration of appropriate technologies such as a statistics software package. Unfortunately, this document fails to recommend any appropriate software or related activities. In addition to finding and learning the software, the teacher is now faced with the task of locating or developing usable activities. In either case, it is a very time consuming and daunting process. Many of these same teachers I previously referred to also reported having little or no formal training with computers or spreadsheet software. Regrettably, this also represents my personal background and experience.

There is a shift in mathematics education away from drill and practice teacher-directed instruction. The teacher's role is now to help direct and focus student activity so those students can construct and build their own meaning and understanding. Increased emphasis is being placed upon:

- actively involving students in constructing and applying mathematical ideas;
- using problem solving as a means as well as a goal of instruction;
- using a variety of instructional formats: small cooperative groups, individual explorations, whole class instruction, and projects;
- using calculators and computers as tools for learning and doing mathematics (NCTM, 1992, p. vi).

The previously identified items and this new shift in mathematics education has led me to develop a series of activities in *Microsoft Excel*. The intent is to provide a set of curriculum specific investigative activities that incorporate the use of spreadsheet technology into the exploration of select statistical topics pertinent to the curriculum.

### Literature Review

#### The Need for Technology Integration

The NCTM states that "the use of technology in instruction should further alter both the teaching and learning of mathematics" (1989, p. 128). For teachers, this would involve using computer software for class demonstrations. For students, it would involve independently exploring additional examples, performing independent investigations, and generating and summarizing data as part of a project (NCTM, 1989, p. 128). Also, because technology is changing mathematics and its uses, the NCTM believes that:

- appropriate calculators should be available to all students at all times;

- a computer should be available in every classroom for demonstration purposes;
- every student should have access to a computer for individual and group work;
- students should learn to use the computer as a tool for processing information and performing calculations to investigate and solve problems (NCTM, 1989, p. 8).

The NCTM continues to emphasize the role of technology in education in the new draft *Principles and Standards Document*. “Mathematics instructional programs should use technology to help students understand mathematics and should prepare them to use mathematics in an increasingly technological world” (NCTM, 1998).

The computer is not intended to replace the teacher. It is a tool that can guide a student through the development of a mathematics topic using activities that have been carefully worked into the school curriculum (Posamentier and Stepelman, 1999, p. 144). Posamentier and Stepelman (1999, p. 144) refer to such activities as computer assisted instruction and offer that it “should supplement, not supplant, the regular instruction.” The NCTM (1998) also states that technological tools “should not be used as replacements for basic understandings and intuitions; rather, they can and should be used to foster those understandings and intuitions.” In the area of statistics, “computers can be used effectively throughout students’ study of data analysis, statistics, and probability to illustrate concepts, perform calculations, create representations, and provide data from simulations” (NCTM, 1998).

There are two relevant areas that provide direction to the project. The first examines the role of spreadsheet use in the study of statistics. The other considers issues pertinent to the learning environment: cooperative learning, constructivism, and reflective practice.

### Why Use a Spreadsheet to Study Statistics?

Spreadsheet software was initially created for the business world. However, Arganbright (1992, p. 226) claims that an electronic spreadsheet is a powerful and creative tool for the study of statistics. He argues that it provides a natural medium through which to examine statistical ideas interactively, implement algorithms, and generate insightful graphic displays. Other benefits he feels that can be achieved include increased student interest and the preparation of students for many careers. Hall (1995, p. 5) claims that using a spreadsheet package not only aids the learning process but also adds to a student's professional competence. The long-term benefits of spreadsheet knowledge to students for statistical analysis are extended by Nash and Quon (1996).

Spreadsheet technology can also serve as a tool to enhance the exploration of statistical concepts. This technology has the capability of handling large samples of data in performing complex calculations and producing professional charts. In my previous instruction of the high school statistics course, students often reported that working with large sets of data was time consuming, monotonous, and even difficult. Forte (1995, p. 208) offers that spreadsheet software enables students to organize, analyze, and compute statistics from data without performing complex manual computations. Forte (1995, p. 208) claims this will allow students to spend more time analyzing and interpreting the generated data and less time learning complex statistical formulas.

Dreyfus, Feinstein, and Talmon (1998, p. 68) also acknowledge that spreadsheet use provides freedom from complex and tedious calculations when they integrated spreadsheets into a biology curriculum. They further concluded that students became

“actively and personally involved in the construction of scientific concepts and concepts related to the empirical and quantitative statistical character of biological sciences” (Dreyfus et al, 1998, p. 68). They also felt that learning to use a spreadsheet placed no extra cognitive load on students. In fact, “in the eyes of an overwhelming majority of the teachers, the technical aspects of the use of the spreadsheet was not regarded as a source of difficulty” (Dreyfus et al, 1998, p. 78). In addition to the relative ease with which students can learn to use a spreadsheet, a spreadsheet offers other features that can help students investigate concepts.

Mitchell (1997, p. 203) developed a problem-oriented approach to statistics education using a computer-based spreadsheet. In support of his choice to use spreadsheets, he offers five reasons. First, a spreadsheet can be used to create multiple representations of statistical measures. Mitchell (1997, p. 203), felt that “from an instructional viewpoint it is quite useful to present both an analytic (or algebraic) and a visual (or geometric) representation simultaneously.” Through the use of well chosen graphs, students can build models which display the analytic and visual representations dynamically. This feature is made possible by the fact that visual representations and calculations are updated automatically as a piece of data changes (Mitchell, 1997, p. 203). Students immediately see the effects of editing existing data or entering new data. This allows them to actively investigate and “reinforce the conceptual underpinnings of a statistical concept” (Mitchell, 1997, p. 204).

A second related item Mitchell (1997, p. 204) describes is “number playgrounds.” Here, the user inputs raw data for a limited number of subjects and the spreadsheet automatically calculates relevant statistical measures. Students then change or play with

data to experiment with statistical concepts. This focuses on students' "conceptual understanding" (Mitchell, 1997, p. 204).

Two other reasons to use a spreadsheet are offered by Mitchell (1997, p. 204) under the headings of "story lines" and "creativity." Together, he is referring to the capabilities of the spreadsheet, which allow for the addition of text, sound, color, and graphics. These features allow the student to create descriptive products that have high visual appeal.

Finally, Mitchell (1997, p. 204) discusses "the power of open ended software programs." He states that "the potential uses for a spreadsheet program are limited only by the skill and patience of the user" (Mitchell, 1997, p. 205). Mitchell argues that together, "these five elements provide a powerful cognitive foundation for building instructional activities aimed at developing a deeper understanding of statistical concepts" (1997, p. 205).

The need for knowledge and skill with using spreadsheet technology is also being documented in the workplace. Perry (1998) conducted a study of employers who advertised for computer related vacancies in five metropolitan newspapers in the southeastern United States during a month in 1996. Employers were asked to examine a list of spreadsheet competencies to determine the degree to which they were needed for their workplace. They rated each spreadsheet competency as 'Essential to Very Essential' for their business (Perry, 1998, p. 37). The NCTM (1998) states that "if schools do have a level of technology equivalent to the level found outside of schools, and if they do not prepare students appropriately with it, then they are placing students at a serious disadvantage." More and more schools today are being outfitted with computer labs and

the latest in software including spreadsheet software. Incorporating this software into classroom instruction is a responsible first step in preparing students to enter the workforce.

#### Cooperative Learning and Reflective Practice

Slavin (as cited in McManus and Gettinger, 1996, p. 13) defines cooperative learning as a process by which students work together in groups to master the material presented by the teacher. Studies by both Slavin and Johnson and Johnson (as cited in Brush, 1997, p. 52) state that for cooperative learning groups to be effective, they must receive instruction and direction in areas they call positive interdependence, individual accountability, group rewards, and group instruction.

According to Johnson, Johnson, and Smith (as cited in Garfield, 1993), cooperative learning is not having the first people in a class to finish their work help the slower students. It is not a group project that one person does and everyone puts his or her names on as contributors. It is not having students sit side-by-side at the same table and talk to each other as they do individual work or assignments. There must be small groups engaged in helping each other gain understanding as they attempt to solve a problem or complete some task. Developing this cooperative environment has been found to have several advantages.

Cooperative learning structures have been linked to increases in academic performance. Dees (as cited in Keeler and Steinhorst, 1995) reported that the use of learning partners and peer monitors resulted in increased performance in problem solving, quizzes, and tests. Studies by Keeler and Steinhorst (1994, 1995) found that students also achieved higher academic scores while becoming more engaged in the

course material. Giraud (1997) also reported higher test scores for students working in randomly assigned cooperative groups but felt this format was especially beneficial to “those least prepared for statistics.” Benefits of this approach are not limited to only academics.

Johnson and Johnson (as cited in Giraud, 1997) advocate cooperative learning for “its effects on achievement and on other attributes that accompany the acquisition of knowledge, including motivation, classroom socialization, the student’s confidence in learning, and attitude toward the subject being learned.” According to Legere (1991, p. 170), the small groups’ exchange of ideas “increased the interest of the participants in addition to promoting critical thinking.” Whicker, Bol, and Nunnery (1997, p. 47) report that most students enjoyed the social aspect of working with and receiving help from other students. Some justification for this is offered by Forte (1995, p. 211) when he claims that students in a group setting are free from the surveillance of the instructor. They are now free to “take risks in ways that minimize anxiety and embarrassment” (Forte, 1995, p. 211). Brush (1997, p. 51) noted improved students’ attitudes toward math and computer math activities for students who used an Integrated Learning System with cooperative learning pairs compared to those who used only an Integrated Learning System.

Working in cooperative groups, students can avail of group interaction and dynamics to construct their own meanings of concepts discussed. As students investigate a concept, “it is important that provisions be made for students to share their experiences, clarify their thinking, generalize their discoveries, and provide convincing explanations” (NCTM, 1992, p. vi). The questions and observations that result from group efforts can

and should be shared with the whole class. This may open up topics for further investigation.

Reflection is often associated with the “looking back” step in Polya’s plan for solving problems: read, plan, solve, and look back (Krulik and Rudnick, 1994). As a facilitator of understanding and knowledge acquisition, many would argue that reflection is the step in the process that holds the most potential. Wheatley (1992) claims that “reflection plays a critically important role in mathematics learning” and offers evidence that “reflection results in greater mathematics achievement” (p. 529). What is involved in reflection?

Krulik and Rudnick (1994) claim that for many teachers, looking back is simply to “determine if the mathematics is correct, see if it makes sense, and check how well it answers the question” (p. 334). This is not the true reflection that leads to individual insights. For Sigel (as cited in Wheatley, 1992), reflection for mathematics learning “is characterized by distancing oneself from the action of doing mathematics” (p. 535). It involves the deliberate investigation of “one’s own action” in solving the problem. Wheatley (1992) summarizes the power of reflection when he states that “persons who reflect have greater control over their thinking and can decide which of several paths to take, rather than simply being in the action” (p. 535).

In adding some extra substance to philosophical approaches to reflection, Krulik and Rudnick (1994) identify five elements of effective reflection. These include :

- 1) test the reasonableness and practicality of the answer,
- 2) write a summary paragraph about the problem and its solution including thought processes and false starts,

- 3) find other solutions,
- 4) change the conditions of the problem,
- 5) extend the problem to a mathematical formula, concept, or generalization (p.334).

Krulik and Rudnick (1994) claim that “engaging in this process not only improved students critical thinking, but found that students are motivated by, and enjoy the challenge of finding alternate solutions, asking what if, and finding a mathematical expression or concept that describes the situation” (p. 338). In thinking about one’s thinking, students clarify their thought processes and focus their direction for future problems. It awakens the student to diversity that exists within a problem and the attempts with which it may be solved. When reflection is done in small groups or in whole class settings, students benefit from other student’s insights and approaches. Such discussions can also be enlightening to teachers and help shape their future instructional activities. Cobb, Boufi, McClain, and Whitenack (1997) speculate that there is a relationship between “reflective discourse and conceptual development in mathematics” (p. 265). For the teacher, “analysis of reflective discourse clarifies how teachers might proactively support their students mathematical development” (Cobb et al, 1997, p. 274). Teachers and students have a role to play in effective reflection.

*The Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989) encourage students “to explore mathematics, to gather evidence and make conjectures, and to reason and communicate mathematically as they discuss and write ideas that use the language of mathematics” (p.vii). These activities change the role of the teacher from that of an information dispenser to that of a facilitator. The emphasis is not

on showing the students the correct way, but on helping them to formulate an acceptable, usable and valid way to manage data. The teacher will need to pay careful attention to students' small group and whole class presentations. Through reflection on these events, the teacher will determine the degree to which students have formed appropriate interpretations of the required concepts.

As part of the teacher's preparation, teachers must recognize that "the only thing that is predictable in teaching is that classroom activities will not go as predicted" (Simon, 1995, p. 133). Careful attention needs to be given to the background of the group and the learning experiences they have encountered. Listening to students will be a key component of assessing their understanding (Simon, 1995, p. 141). When the students are learning the mathematics, the teacher should be learning about mathematics, learning, teaching, and the mathematical thinking of his or her students (Simon, 1995, p. 141). Through such attention and reflection, the teacher prepares to direct discussions that help direct students to appropriate constructions.

#### Overview of Project

##### Description

Students today live in an ever-changing technological society. It is no longer sufficient for them to leave school simply being able to read, write, and do arithmetic. The NCTM (1998) claim that a legitimate need has been identified for the responsible use of technology with the aim of enriching student learning. Traditional didactic methods and chalk and talk lectures can no longer stand as the only or even primary mediums of instruction. Students need to be free to explore and construct their own meanings from the learning environment. This environment needs to include curriculum specific

activities and the appropriate technologies to allow them to investigate and meet the prescribed course objectives.

Among the objectives of the Data Management unit of Course 1, Level Y, *Teacher's Pilot Draft Guidelines: Atlantic Canada Mathematics Curriculum*, students are required to:

1. Calculate/determine manually and with technology the mean, median, mode, range, and standard deviation of a set of data.
2. Use the mean, median, mode, range, and standard deviation to describe a set of data.
3. Produce/create manually and with technology tables, scatter plots, stem-and-leaf plots, and histograms for the presentation and investigation of data.
4. Use interpolation, extrapolation, and the equation to predict and solve problems.
5. Use the regression analysis capabilities of technology to determine the line of best fit.
6. Develop an intuitive understanding of correlation and the correlation coefficient.

Through a series of six activities, this project provides students with opportunities to learn and investigate the previously mentioned concepts of statistics as they learn various features of *Microsoft Excel 97*. Students will use *Excel* as a tool to free them from time consuming and tedious calculations. This will enable them to have more time to explore and conjecture based on their observations. *Excel* will allow them to produce professional looking charts and tables so that they can illustrate and support or disprove their conjectures.

#### Classroom Climate

The activities are created with the assumption that students will have access to a lab containing appropriate technology. The nature of the teaching/learning environment

may vary depending upon laboratory constraints. It is suggested that students engage in each activity through a cooperative learning atmosphere developed by the teacher. For example, in my teaching environment, each computer lab is equipped with 20 computers and there are 30 to 35 students in a typical class. A computer would be assigned to each group of 2 or 3 students with extra computers available in the case of breakdowns. Each group member would alternate turns entering data and commands into the computer. Students would discuss the relevant questions at the end of each activity in an attempt to arrive at solutions agreeable to both members. At the conclusion of each activity, each group would then be invited to share/present their findings and insights with the whole group for further discussion and analysis.

These activities are not intended to entirely replace regular classroom instruction. Prior to the laboratory experience, some traditional instruction relating to the relevant statistics concepts may be required. Instead, the laboratory activities are intended to be used more like science labs. They provide a medium for students to learn, explore, reinforce, and apply statistics concepts.

#### Note to the Teacher

The activities that follow are intended to provide high school teachers with easily accessible materials that will make the integration of technology less threatening. This resource features six activities that appear in teacher and student versions. Both copies may be reproduced for classroom use provided that acknowledgement is given to the author.

The first set of activities encountered is the teacher's versions. Each activity is preceded by an overview of the activity and the objectives it intends to address. The

actual activities that follow this include additional details that are intended to help troubleshoot some of the technical problems that may arise. Obviously, it is impossible to predict all such problems.

An accompanying set of student activity sheets is included in Appendix B. The student sheets address all the same ideas as the teacher sheets but some of the extra explanation is omitted. As both sets of activities progress, some previously addressed *Excel* commands and procedures are treated as assumed knowledge. If the user decides to skip an activity, they should be careful to ensure that the flow of the activity would not be interrupted by a lack of prerequisite *Excel* knowledge.

#### Sequencing the Activities

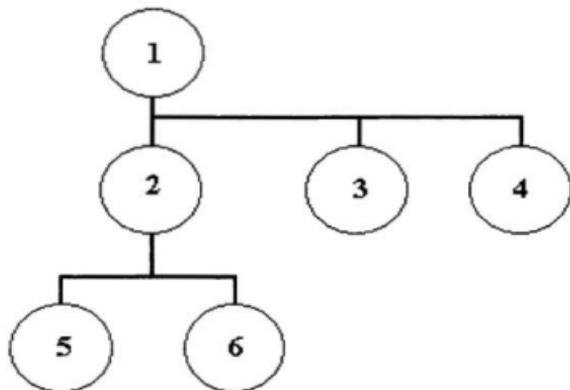
Teachers and students should read through the **Background to Excel** (see Appendix A) before commencing the activities. The background reading is best done in the lab where the user can open *Excel* and view the items discussed. The teacher may choose to guide the students through the introductory material.

The activities that make up this project are designed to address the concepts in the order that they occur in the course objectives. Many of the activities are based on the data collected in the first activity with previously introduced *Excel* commands treated as assumed knowledge in latter activities. It is not necessary to do all 6 activities nor do them in order. The following is a summary of relationships between the activities:

- Activity 1 is a prerequisite to all of the other activities.
- Activity 2 uses the Chart Wizard and is a prerequisite to activities 5 and 6 which are independent of one another.

- Activities 1, 2, and 3 presently use the same set of data for all operations with activity 6 revisiting that data for further analysis.
- Activities 3 and 4 are independent activities.

The following flow chart summarizes the suggested sequence of the activities by number:



The degree to which these activities will meet the needs of a teacher will vary based on his/her own strengths and those of the students. Individual teachers may feel that other teaching methodologies are required to meet the objectives of the curriculum. For this reason, the instructions are intended only to help focus student direction. Although not specifically addressed in the activities that follow, homework complements class work and should be a regular part of this unit. The prescribed textbook can be used alone or in conjunction with *Excel* to reinforce the objectives covered in this unit.

No activity should require more than two hours of lab sessions, however the quality of computer hardware may influence this timing. The activities of this project are developed using *Excel 97* in *Windows 95*. In addition to ensuring the appropriate software

is available, the teacher should check the **Tools** menu of *Excel* to ensure **Statistical Analysis** is installed as an option. This particular function is optional and often left out when *Microsoft Office* is installed. Check with your local technology expert, especially if the computers are on a network.

### Evaluation Plan

At the end of every unit, the students will need to undergo some summative evaluation. At this point, a knowledge of *Excel* is not listed as a performance objective of the Data Management unit. However, since students are required to integrate technology into the analysis of data, some items that test such knowledge may be legitimate.

This set of activities is intended to be a resource to teachers. If it is to be accepted as appropriate, it should be subject to some form of formative evaluation. Through reflection on the use of these activities, the user will be able to shape these activities into a better resource or justify a total rejection of them. Some questions that may help direct this reflective evaluation process are offered here:

- Were the objectives/activities presented to students in a suitable sequence?
- Were the course objectives met?
- Was student performance on testing improved, decreased, or the same?
- Was the instruction suited to the level of the students?
- How were individual differences accommodated?
- Could the course objectives be more appropriately met through other technologies?
- Was there sufficient computers to conduct each activity?
- Was a reasonable amount of time allocated for completing the unit?
- Was there ample class time to complete each activity?

- Should some activities be omitted?
- Which activities best met the educational objectives of the course?
- Was the method of testing students appropriate?

These and many more questions should be carefully asked and answered in the search for an appropriate method to integrate technology that enhances students learning and development.

## Data Management: A Set of Activities in *Excel*

## Overview of Activity 1

In this activity, students will use a survey to gather data on student age, grade level, hours a week studying math, math average, height, and ability to shoot free throws in basketball. This activity will introduce students to entering data into a spreadsheet and explore some of its calculation capabilities as they apply to the study of statistics. This data will be used for future activities and will need to be saved.

### Objectives

#### Statistical topics.

Students will collect and organize data. They will analyze the data using measures of central tendency (mean, median, mode), maximum/minimum values, and range.

#### Mathematical thinking.

Students will use *Excel* to analyze data and arrive at definitions for each of the mathematics terms above and outline how each is found. They will also be encouraged to make inferences and conjectures based on these calculations.

#### Technology.

Students will learn how to enter text and data into cells, save, and print the results. Using *Excel*, they will perform basic calculations and functions such as **MAXIMUM**, **MINIMUM**, **RANGE**, **AVERAGE** (mean), **MEDIAN**, and **MODE**. The **copy** and **fill** options will be used to apply formulas and functions to adjacent cells. The **sort** function will also be employed to arrange data to assist analysis.

## Activity 1

Part 1: Preparing to collect and save data

Present students with the following situation:

Your consulting firm has been asked to gather the following class data for future analysis: name, age, grade level, hours spent studying math (week), math average, height (cm), and basketball free throw ability. Answer all questions honestly and take steps to ensure accurate measurements. Each student attempts ten baskets and you record the number of successes.

Distribute a class list of student names and open *Excel* from the start menu as before. For consistency, move to cell **B1** and type the title STUDENT PROFILE. Notice that the words also appear in the block to the right of the **Name Box**. This box is called the **Formula Bar**. Pressing Enter or one of the arrow keys will now assign these words to the cell **B1**. To edit this cell you can move to it, retype the entry and press Enter. This replaces the last entry. If the mistake is minor, then double click on the cell. Now you can edit the typing and press Enter to have the change accepted *Excel*, like word processors, also has a spell checker for tidying most of this up at the end.

---

\*\*\* Activity 1 is an adaptation of an activity from an article entitled "Exploring Mean, Median, and Mode with a spreadsheet" by Wilson and Krapel (1995).

Now let's name the columns. In cell **A4** type **Name**. Some headings may be two words so you may like to enter the words in two rows as below in **B3** and **B4** or **C3** and **C4** for example.

	A	B	C	D	E	F	G
1		STUDENT PROFILE					
2							
3			Present	Hrs. Study	Math	Height	Baskets
4	NAME	Age	Grade	Math(wk)	Average	(cm)	(/10)
5	John Doe						

Keep the headings short but descriptive. If names are too long for the formatted cell, do not worry. To fix this, begin by moving the pointer to the vertical line at the right of the column letter. When the plus sign changes to a pair of vertical parallel lines, double click it. The column is now formatted to the width of the longest entry. Proceed to enter the other column names. Enter the student names under **NAME** in cells **A5**, **A6**, etc.

When you have completed this exercise, save the work for future reference. To help organize your work, you should create a folder to store all related files. Click the disk in the upper left corner and you will get a '**Save as**' message. Click the third folder to the right to create a new folder to store your *Excel* files. In the **New Folder** screen that appears type *Excel Files* in the **Name** box and click **OK**. A new folder will appear. Click the folder so it appears in the **Save in** box. Name your new workbook by typing **Student Profile** in the **File name** box at the bottom and click **Save**. The workbook is now saved in your *Excel Files* folder as **Student Profile**. You should save your work regularly.

Now to see what your file will look like, go to **File** and select **Print Preview**. Notice that there are no lines to work with and you may be missing a column. Select **Close**. To get a grid, select the range in which you want the lines beginning with **A3**.

Select **Format**, **Cells**, and **Border**. Click **Outline, Inside**, and select **OK**. Preview your document and observe the changes. Were you missing a column? Select **C**lose.

If you were missing a column on this sheet it means that it would be printed on the next page. This can sometimes be avoided by orienting the paper sideways using the landscape feature. Select **File**, **Page Setup**, **Page**, click **Landscape**, and **OK**. This turns the paper on its side so that more columns can be displayed. Preview your document and observe the changes again.

Select Print and note your option to print the entire document or just select pages. For now, select **OK** and you will have a sheet with which to collect data. Remember to save before you exit. To exit the workbook, select **File**, **Close**, and usually hit **Yes** if you made changes that you want to save. Then click the  in the upper right corner of the screen to close *Excel*.

Congratulations, you have just opened, formatted, saved, and printed a potential worksheet. For the beginner, this is no small task. In the next phase of the activity, students will gather data and begin the analysis of it.

#### Part 2: Collecting and entering data and formulas

Divide the class into groups of three and have them collect and record the data in the table they created. A team leader will need to be selected or nominated to direct the activities of each group. When this process is complete, direct students to open their *Excel* workbook and begin entering their data. For consistency, all students will need the same class information so they will need to share their data sheets until they each have the required data.

Next, skip a row and enter the statistical terms to be investigated in column A as below:

	A	B	C	D	E
30	Maximum				
31	Minimum				
32	Range				
33	Mean(Avg.)				
34	Median				
35	Mode				
36					

In the cell adjacent to the word **Maximum** enter `=maxa(B5:B20)` and press Enter. This assumes that **B5** is the first element in the range and **B20** is the last. The equals sign, "=", tells *Excel* that a formula or predefined task like **maxa** will follow. After the task, open brackets to enter the range of cells that contain the numerical data separated by a colon and close the brackets. When you open the brackets you can quickly define the range by highlighting it with the mouse and closing the bracket at the end.

In the cell adjacent to **Minimum** enter `=mina(B5:B20)` and press Enter. In the cell adjacent to **Range** enter `=B30-B31` and press Enter. In the cell adjacent to **Mean** enter `=average(B5:B20)` and press Enter. In the cell adjacent to **Median** enter `=median(B5:B20)` and press Enter. In the cell adjacent to **Mode** enter `=mode(B5:B20)` and press Enter. Carefully study the numbers that appear.

You will also want to calculate these values for the other columns of data. We can take advantage of *Excel's* copy command to save some typing here. First, highlight the range of cells that we just entered formulas into. Next, click the copy icon at the top of the page next to the scissors and a dashed border surrounds the range of cells. Highlight

the entire range of cells that will hold the information and click the paste icon (next to copy). *Excel* performs the same calculations on each of the other columns.

Add some customization to your worksheet by double clicking the **Sheet 1** tab at the bottom of the sheet. The word **Sheet 1** will be highlighted. Simply type a new title for the page like **Initial Data** and hit **Enter**. Sheet 1 is now called Initial Data. Save this file replacing the previous worksheet. Print the sheet and submit it with your solutions to the questions in part 3 of this activity.

Next, make a copy of the original data. Highlight all the cells used in the **Initial Data** sheet and click the **copy** icon. You can also select **copy** by clicking the right mouse button and selecting it from the menu that appears. Select the **Sheet 2** tab at the bottom of the page and click **paste**. All the data is now transferred to **Sheet 2**. Rename this sheet as 'Working Data'. Feel free to alter this data as you answer the questions in part 3.

At times, it may be easier to analyze the data by sorting it using some criteria. Highlight the entire range of data below the column headers (do not include the calculations you made or the headers). Select **Data** and **Sort**. In the **Sort** dialogue box click the down arrow in the first **Sort by** section. Each of the column headers will be there. Select **Age** and **OK**. Sort again and as often as you please but remember to highlight the entire range so that the overall arrangement of data is preserved.

### Part 3: Analyzing the data and making conclusions

1. Based on the data, write a definition for maximum, minimum, and range.
2. What relationship exists between maximum, minimum, and mean (avg.) in each of the columns?

3. What will happen to the maximum, minimum, and mean (avg.) if one of the entries is changed to zero? What happens if one value is made extremely big?
  4. Look at the sorted data values in relation to the value that represents the mode. Write a definition for mode.
  5. Look at the sorted data values in relation to the value that represents the median. Write a definition for median.
  6. Change a data value to zero or to an extremely large number. What effect does this have on the mode and the median? Is this consistent with your definition?
  7. Experiment and change some data elements. Explain how the median is similar to but different from the mean.
  8. Which single descriptive statistic - maximum, minimum, range, mean, median, or mode - best describes each of the categories?
  9. If students had been asked to name their favorite sport, could either of mean, median, or mode be used to describe the resulting data? Why?
- \*\*\*\*\* Remember to submit a copy of your Initial Data sheet with your work.

## Overview of Activity 2

This activity uses data gathered in the previous activity. It will introduce students to the graphing capabilities of a spreadsheet. They will use these graphs to explore relationships among pairs of items.

### Objectives

#### Statistical topics.

Students will create **line graphs**, **bar graphs**, and **scatter plots**. They will use these graphs and the statistical calculations done in the previous activity to discuss and make conjectures on possible relationships among data. Sample size and population will also be discussed.

#### Mathematical thinking.

The students will analyze the graphs and the statistical calculations. They will make inferences and conjectures on potential relationships or trends between pairs of items.

#### Technology.

Students will learn how to create **line graphs**, **bar graphs**, and **scatter plots** using the Chart Wizard. They will add titles, format titles, define axis, change scales on axis, and position and size charts in the worksheet. Students will also copy the chart into a brief form letter in *Microsoft Word*.

## Activity 2

Part 1: Creating and editing a chart

Present students with the following situation:

Is there a relationship between study time and math grades? Display these relationships in a separate bar graph, line graph, and XY scatter graph. Your company would like you to summarize your observations and include them with the charts in a brief written report.

Open *Excel* and open the file **Student Profile**. Now copy the original data to **Sheet 3**. Click the **Initial Data** sheet tab and select entire range of data. Select **copy**; click the **Sheet 3** tab at the bottom of the page and click **paste**.

Create the chart.

Use the **Chart Wizard** to create the graphs. This function leads the user through a simply four-step procedure. Begin by highlighting the range of cells to be included in the chart. To graph Hours of Study versus Math Average, highlight the entire range of data including the column titles in columns **D** and **E**. When the ranges are not adjacent, highlight one range, press and hold the control key **Ctrl**, and highlight the other range.

With the range selected, click the **Chart Wizard** icon. If the Office Assistant offers help, click **NO**. The **Chart Wizard** dialogue box indicates that this is 'step 1 of 4 - chart type'. Click **Column** from the list that appears and the types of column charts will appear to the right. Select the chart type 'clustered column, compares values across categories' and then **Next>**.

In step 2, you are asked for the chart source data. The range you selected is already in the blank opposite **Data range**. Ensure that **Columns** is selected in the 'Series in' section and click **Next>**.

Step 3 allows for some customizing of the chart. Select the **Titles** tab from the above list. For the title enter 'Math Avg. versus Hours Study (wk)'. Similarly, label the x-axis 'Hours Study(wk)', the y-axis 'Math Avg.', and click **Next>**.

The final step simply asks if you desire to have the chart created on its own page or placed in the present sheet. Select '**As object in**' and **Finish**. The chart will appear on top of the data in your worksheet. Click and hold the mouse button on a white part of the chart. When the pointer changes to a four-directional arrow, you can drag the chart to another part of the worksheet. Place it directly below the data by dragging the border.

#### Edit the chart.

To edit the chart, click the white area of the chart window and the chart Edit window is activated. You may then click on any object in the chart to activate it. When activated, the object will have handles (small black boxes) around it. This allows the user to make any desired changes.

Click on the upper right handle of the chart and drag the border so that it rests in the upper right cell one row below the data. Next, click on the lower left handle and drag the border to make the chart six rows and six columns larger.

We can also change fonts and move text in the chart. Activate the Edit window and click the title (a frame will surround it). Select **Font Size** from the menu above and choose 14. Next click on the frame and move it to the center. This can also be done with the labels on the axis.

The scale that appeared on the axis can also be modified. Click on the vertical axis and the handles appear again. Next, click the right mouse button and select **Format axis**. Select the **Scale** tab. Turn off the check mark in **Major unit** and type 10 in the blank. Turn off the check mark in **Minor unit** and type 5 in the blank. Select **OK**.

You may also want to highlight the axis and add gridlines to help the reader interpret the results. Activate the vertical axis line and click the right mouse button. Select **Format** and the **Patterns** tab. Click **Axis, Custom**, select **Weight**, click the thick line, and **OK**. Repeat this process for the horizontal axis. To get gridlines, click the white area of the chart window to activate it. Click the right mouse button and select **Chart Options**. Select the **Gridlines** tab. The horizontal axis is **Category (X) axis**. Select only the **Minor gridlines** here and in the **Value (Y) axis**. Select **OK** and close the Edit window by clicking outside the chart. Save and print your work.

Change some data values. Notice that the chart is automatically updated to show the change. When you exit, answer **NO** when asked if you want to save changes to the worksheet. This will preserve the original data for the next activity.

### Part 2: Creating line graphs and scatter plots

Follow the set of instructions from Part 1 to create a Line chart of the type 'Line with markers at each data value' and place it under the previous one. Similarly, create a XY (Scatter) chart of the type 'Scatter, compares pairs of values'. Use only the formatting you feel is necessary to clearly display the data.

Copy your charts to a new page in the workbook. Arrange the charts so that they can all be printed on the same page (fit them in between the dashed lines). You will have to fit the charts into this space. Save and print your work.

### Part 3: Interpreting and presenting your findings

The following questions will be answered in paragraph form and the solutions typed in *Microsoft Word*. This is a word processing program that works much like *Excel* for formatting, saving, and printing a document. As part of your response to question number three, you will be asked to copy and paste a chart into your work.

To copy the chart, minimize *Microsoft Word* by clicking the first of three boxes in the upper right corner with a flat bar in it. Open your *Excel* file as usual and click the white part of the chart you want. Select **Copy** and a border appears around the chart. Close *Excel* and reactivate your *Microsoft Word* document by clicking on it in the status bar by the start menu at the bottom of the screen. Move to the location that you want the chart and select paste. Do not worry about the appearance or formatting of your work in *Microsoft Word*. Finish typing your answers, then save and print your work. The questions are as follows:

1. Describe any relationships you feel may exist between success in mathematics and hours of study. Do this without looking at the data you collected.
2. In looking at the raw data alone (not the charts), describe what you observed that supported or went against your ideas of a relationship?
3. Of the three types of charts explored, which do you feel best displays and supports your ideas? Why?
4. The sample of data came from your class. Could it be used to describe all the students in the school, the province, or Canada? Support your views in each case with some logical argument.

### Overview of Activity 3

In this activity, students will use data gathered on student math averages to create a stem-and-leaf plot. They will also create histograms of the respective data sets for comparison to stem-and-leaf plots. Charts will also be discussed in relation to calculations involving mean and standard deviation.

#### Objectives

##### Statistical topics.

Students will use a stem-and-leaf plot to organize data and create a histogram. They will analyze the data with measures of central tendency (mean, median, and mode).

##### Mathematical thinking.

Students will create a stem-and-leaf plot to organize data in bins (groups) by order of size. They will investigate how this plot can be used to arrive at numerical values for median and mode. Students will compare this chart to a histogram.

##### Technology

Students will learn how to insert new pages into a workbook, change cell widths on a sheet, and format a range of cells. They will use the **cut**, **paste**, and **sort** functions to organize the stem-and-leaf plot for comparison to a histogram. The Data Analysis tool bar will be used to create a histogram.

## Activity 3

Part 1: Creating the Stem-and-Leaf Plot

Open your *Excel* workbook. To enter a new sheet, select **Insert** and **Worksheet**. Add a couple of sheets while you are here. Click the Sheet 4 tab to begin working on a new sheet. Change the name to Stem-and-Leaf plot.

A stem-and-leaf plot allows for organizing data into classes for analysis while preserving the original data values. In this activity, we will use the math average data collected for Activity 1. The data are marks out of 100 so the stem or fixed items in the data are the tens digits Type **Leaves** in cell **A10** and type **Stems** in **A11**. In columns **B11**, **C11**, **D11**, ... , **L11**, enter 0, 1, 2, ... , 10 as below.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
:												
:												
9												
10	Leaves											
11	Stems	0	1	2	3	4	5	6	7	8	9	10
12												

The stem labels represent the intervals like 0 to 9, 10 to 19, and 20 to 29 respectively. To separate **Leaves** and **Stems** as above, select the range **A11** to **L11** for formatting. Select **Format**, **Cells**, **Border**, and double line in **Line Style**. Click only the top border and **OK**.

To record a data value, we now enter the ones digit above the appropriate tens digit. For example, if 34 were the first data value, enter 4 above the stem 3. For 30, you place a 0 above the 4 in the 3 stem. An 85 would result in a 5 over the 8. Continue to enter the data in this manner.

### Part 2: Formatting the stem-and-leaf plot

Starting in rows 10 and 11 should allow you to do most stem-and-leaf plots with a small data set. If you need more rows, simply go to row 1 or any place above the highest entry and select **Insert** and **Rows**.

To close up the space between the columns, move the mouse pointer to the line to the left of each column letter. When the pointer changes to a set of vertical parallel lines, click it. The column is now formatted to the width of the longest entry in the column.

It is often easier to read the chart when a vertical line separates each column. Highlight the range of cells from **B11** to **L11** and up to highest row used. Select **Format**, **Cells**, **Border**, and a solid single line in **Line Style**. In the **Border** section of this dialogue box, select each of the vertical line options and **OK**.

Select **File** and **Print Preview** to view the results. Close. In cell **A1**, enter the title "Stem-and-leaf Chart for Math Averages". Select any cell in the upper right corner of the page and enter your group number and the names of those in your group in the cells below.

### Part 3: Drawing a Histogram

*Excel* will automatically sort data into several predefined ranges to create a histogram. To control the bin widths for the sorting of data, you must plan ahead. For now, sort the data values in the intervals (bins) of 0 to 9, 10 to 19, 20 to 29, ... , 80 to 89, and 90 to 100. Return to your Initial Data sheet. Go to column **H** and enter the heading **Bin Width** in cell **H1**. To get the desired class widths, enter 9, 19, 29, 39, 49, 59, 69, 79, 89 below in the respective column **H** cells.

The histogram is not contained in the Chart Wizard. Instead, you must select **Tools, Data Analysis, Histogram**, and **OK**. In the **Histogram** dialogue box, click in the line for **Input Range**. Highlight the set of data values in the Math Average column and the range appears in the **Input Range**. You can type this in manually if you prefer. Next, click in the **Bin Range** line and highlight cells containing the range values. Before moving to the output options, ensure the Labels box is not activated with a checkmark. *Excel* will generate appropriate labels in the output table. In the **Output options**, select **New Worksheet Ply** and ensure only **Chart Output** is activated by a checkmark. The histogram and output table will be created on a new sheet in the workbook. Rename this sheet as Histogram 1.

#### Part 4: Formatting the Histogram

The histogram can be customized in much the same way as charts created with the Chart Wizard. Click the white area of the chart window to activate the chart Edit window. Drag the lower right handle to expand the histogram over 2 more columns and 6 rows.

Note the spaces between the bars. What chart does it look like? To remove these spaces, right click on any vertical bar. Select **Format Data Series, Options**, set the **Gap Width** to zero, and click **OK**. It now looks like a true histogram.

For some final touches, recall some work from the previous activity to personalize the histogram. Change the title of the chart from "Histogram" to "Histogram for Math Averages." Change the label on the horizontal axis to something like "Upper Bin Boundary." Make the horizontal and vertical axis a heavier line. Then, move the legend. Right click on Frequency, select **Format Legend, Placement**, click **Bottom**, and click **OK**. Now left click and hold to drag it to the lower left corner.

Part 5: Comparing and analyzing the charts

Copy the stem-and-leaf plot to the same page as the histogram for comparison. Go to the page called Stem-and-Leaf plot. Highlight the required range and select copy. Return to the Histogram 1 sheet and select a cell below the histogram. Select paste. Print this sheet.

Answer each of the following questions:

- 1.(a) Explain how a stem-and-leaf plot facilitates finding the mode of a set of data.
- (b) Explain how a stem-and-leaf plot facilitates finding the median of a set of data.
- (c) Explain how a stem-and-leaf plot can be used to quickly add the data for finding the mean.
2. Of mean, median, and mode, which best represents this set of data? Explain why.
3. What does a stem-and-leaf plot show about the data that measures of central tendency do not?
4. Compare the stem-and-leaf plot you created to the histogram. What similarities do you notice and what information is lost in the histogram?
- 5.(a) What is the total area under the individual bars that make up the histogram?
- (b) What is the area under the individual bar for the range 70 to 79?
- (c) What percent of the total area is contained in this interval?
- (d) Of the total number of values, what percent fall in the range 70 to 79?
- (e) For any interval containing values, find the area and repeat (c) and (d).
- (f) Is there a relationship between the area of a region in the histogram and the frequency of values in the corresponding range? Explain your answer.

- 6.(a) What intervals would you need to have five equal width bins?
- (b) Go back to the Initial Data sheet and adjust (edit) the **Bin Width** column to obtain these bins. Create a new histogram on a new sheet as before. Name the sheet Histogram 2 and print it.
- (c) Describe how this compares to the previous histogram. Which offers a better picture of the data? Why?
- (d) Predict what would happen if you used a bin width of five.

## Overview of Activity 4

In this activity, students will enter a set of data and use it to investigate how standard deviation is calculated. They will then explore how standard the deviation measures the spread of a set of data.

### Objectives

#### Statistical topics.

Students will learn the steps involved in calculating the standard deviation. They explore the relationship mean and standard deviation. They will discuss outliers and the spread of the data.

#### Mathematical thinking.

Students will follow the process involved in the calculation of the standard deviation. They will introduce outliers and manipulate the data to investigate relationships between mean and standard deviation. Students will develop an understanding of how standard deviation measures the degree to which data values cluster around the mean.

#### Technology

Students will enter text and data into cells. They will name a range and use it in calculations involving formulas. Preprogrammed functions for **Mean** and **Standard Deviation** will also be used.

## Activity 4

Part 1: Preparing to investigate standard deviation

Present students with the following situation:

Open a new workbook in *Excel* and set up a spreadsheet to match the format below:

	A	B	C	D	E
1		Calculating Standard Deviation			
2					
3		Data	Deviation	Squared Deviation	
4		X	X-Mean	$(X-\text{Mean})^2$	
5		10			
6		20			
7		30			
8		40			
9		50			
10	Sum				Variance
11	Mean				
12	Standard				Standard
13	Deviation				Deviation

The best known method of finding the standard deviation begins by calculating the mean. Before we can find the mean (average), we need to find the sum of the five data elements. To find the sum, begin by activating cell **B10**. Next, select **Insert, Function**, and select **Statistical**. Select **SUM** from the **Function name** list and click **OK**. In the Number 1 box, enter the range **B5:B9** manually or by highlighting it with the mouse. If the resulting

---

\*\*\* Activity 4 is an adaptation of an activity entitled "Discovering Important Statistical Concepts using Spreadsheets" by Hunt and Tyrrell, 1995 at <http://www.mis.coventry.ac.uk/research/discus/discus.home.html>

dialogue box covers your data, click and hold on a gray part of the box and move it as required. Repeat this process to find the mean and standard deviation in adjacent cells **B11** and **B12**. These are found in **Statistical** functions as **AVERAGE** and **STDEVP** respectively. Check that the value for the mean concurs with your definition of average.

You now know the value for standard deviation without any real labor. Continue the process in its calculation in hopes of receiving enlightenment towards what it measures. Defining a cell or range of cells with a name is often helpful for calculations. The name can then be used in a formula and this eliminates always having to enter the range.

To name the cell **B11** as "**mean**", click the cell (or range). Select the down arrow in the name box and type "**mean**" and press Enter.

Calculate the deviations from the mean. Click cell **C5** and type **=B5-mean** and Enter. Note the value. Copy this formula to **C6:C9**. Some of the values will be positive and some will be negative. Copy the formulas for **Sum** and **Mean** to their adjacent cells and note the values. They should always be zero. If you get a value like **7.11E-15** in this cell, it is giving you  $7.11 \times 10^{-15}$  or 0.000 000 000 000 007 11. The value exists because of rounding by the computer for values that do not divide evenly. True, the value is not zero. However, it is close enough to zero that students will likely accept the property that the sum of the deviations from the mean is always zero. This helps with the justification of the next step.

Explain to students that to get around a sum of zero so that a usable mean can be found, we square the deviations. Click cell **D5** and type **=(B5-mean)^2** and Enter. Note the value. Copy this formula to **D6:D9** and copy the formulas for **Sum** and **Mean** to their

adjacent cells in **D10** and **D11**. The mean of the squared deviations is called **variance**. Standard deviation is the square root of **variance** or square root of the mean of the squared deviations. Explain that this kind of reverses the previous squaring. Name cell **D11** “var” for variance as before. In cell **D12** type  $=(\text{var})^{0.5}$ . The value should be the same as in **B12**. You are now ready to change values and investigate how these changes affect mean and standard deviation.

### Part 2: Investigating standard deviation

Change the data values as required to complete each of the following questions:

1. (a) What effect does substantially decreasing a value have on the mean and the standard deviation?
- (b) What effect does substantially increasing a value have on the mean and the standard deviation?
2. (a) Is the sum of the deviations from the mean always zero?
- (b) Can all the deviations from the mean be positive? Why?
- (c) Can all but one of the deviations be positive? Give an example.
- (d) Describe what happens if all the original integers are negative.
3. (a) Find a set of five values so that the mean is larger than the standard deviation.
- (b) Find a set of five values so that the standard deviation is larger than the mean.
4. For a set of five numbers record the mean and standard deviation. Add ten to each number. What happens to the mean and the standard deviation? Try another number. Is there a general rule to describe the mean and the standard deviation for addition of a constant to each data element?

5. Record the mean and standard deviation for another set of five numbers. Multiply each number by ten. What happens to the mean and the standard deviation? Try another number. Is there a general rule to describe the mean and the standard deviation when each data element is multiplied by a constant?
6. Open the initial data sheet from activity 1. Copy the data to Sheet 3 and name the sheet **Standard Deviation**. Below the names in column A, label rows for **Mean** and **Standard Deviation**. In the adjacent cells, use the predefined *Excel* functions to calculate these values for each column. Use the raw data, mean, and the value for standard deviation to describe how the data are dispersed in relation to the mean for each column.

## Overview of Activity 5

In this activity, students will enter and analyze data gathered from an accelerated motion experiment. Students will draw charts and perform calculations to find the equation of the line of best fit. They will also interpolate and extrapolate information from the graph and data in the chart.

### Objectives

#### Statistical topics.

Students will square data values and use the data to investigate an equation for the graph. They will determine the parameters (slope and intercept) for the equation of a line. Interpolation and extrapolation of information will also be investigated.

#### Mathematical thinking.

Students will program formulas and perform calculations on data to find the equation for the line of best fit. The graph, line of best fit, and data will be used to interpolate and extrapolate information. Students will also manipulate the equation of the function.

#### Technology

Students enter data and text into cells, save, and print the results. They will enter formulas and copy them to other cells. They will apply the statistical functions of **slope** and **intercept**. The Chart Wizard will be used to create a chart. Charts will be edited and modified. The line of best fit will be inserted.

## Activity 5

Part 1: Enter the data and create a chart

Present students with the following situation:

A cart was rolled down an inclined plane for a Motion-Time experiment. The data below was collected. Enter the data into a spreadsheet. Use the given steps to find the line of best fit for the data, the equation of the line of best fit, and the equation of the actual function. Use these items to extrapolate and interpolate the information indicated.

## ACCELERATED MOTION EXPERIMENT

Distance versus Time Data

DISTANCE(cm)	TIME(s)
0	0
2.1	1
8	2
16.2	3
35	4
42.4	5
78.3	6
94.6	7

For consistency with these instructions, begin entering the above data starting in cell **A1** with "ACCELERATED MOTION EXPERIMENT". Do not skip any rows. Your final data entries should be in cells **A11** and **B11**. Activate the **Chart Wizard**, select the **XY (Scatter)** chart using the 'Scatter with data points connected by smoothed lines' type, and click **Next>**. For **Data Range**, highlight the data in cells **A4** to **A11** and click **Next>**. Click the **Chart Title** and type 'ACCELERATED MOTION'. Click and enter **Time(s)** and **Distance(cm)** in **Value(X)** and **Value (Y)** respectively. Hit **Next>**. Select **As object in** and click **Finish**. Note were the values start on the x-axis.

We now need to input the x-values. Click on the graph to activate the edit window and click on any data value (it will change color). Right click the mouse on a data value, select **Source Data**, and the **Series** tab. Now click in the X-Values entry box, highlight (enter) the range **B4:B11**, and click OK. The true data values are reflected on the graph.

To make the graph more readable, we will do some editing as in previous activities. First, change the scale on the x-axis so that **Auto Maximum** is 7 and the **Major Unit** is 1. Highlight both of the axes for emphasis. To add gridlines to the graph, begin by clicking on the graph background to activate it. Right click the mouse, select **Chart Options**, and the **Gridlines** tab. Select **Major Gridlines** for both the X and Y axes and click **OK**. Size and move the graph so that the upper right corner is in cell **A15** to prepare to analyze the data.

#### Part 2: Analyze the data and redo the chart

When analyzing experimental data, we would like to have a straight-line graph. The graph of the data appears to be a segment of a parabola. This indicates that the x-axis data should be raised to a power. Square the time in column C to try to get a straight line.

Enter the header "**Time<sup>2</sup>**" in column C3. Type **Time2** in the cell. Highlight the 2, select **Format, Cells, Font**, click **Superscript**, and **OK**. In cell C4, type **=B4^2**. This formula may then be copied to the range **C5:C11** as in a previous activity. Another quick way to copy this is to first select cell C4. Then, click on the small square box in the bottom right corner of the cell and drag it down to cell C11. You are now ready to draw a new graph for Distance versus Time Squared.

Begin by copying the old chart. Click the chart, select **Copy**, and move to cell A35 and select **Paste**. A duplicate chart will appear for editing.

To enter the time squared data, click in the chart to activate the edit window. Activate the data points and right click a point. Select **Source Data** and click in the **X Values** line. Delete the contents of this line. Enter (highlight) the range **C4:C11** and click **OK**. We now have a straight line, but the scale on the x-axis only goes to 7.

For some final touches, edit the x-axis to change **Maximum** to 50 and the **Major Unit** to 10. Then, we need to change Time to Time<sup>2</sup> in the title and on the x-axis. Double click the cursor after the 'e' in time so it appears as a flashing vertical bar. Type a '2' and then highlight it. Select **Format, Selected Axis Title, Superscript**, and **OK**. Repeat this in the title and click outside the chart to turn off the edit box. Save the work.

### Part 3: Getting the equations

To get the equation of the function for graph one, we must first determine the equation of the line in the Distance versus Time<sup>2</sup> graph. That is, we must get the slope and y-intercept (parameters) for the equation of the line.

The equation of a line is traditionally written as  $Y=mX+b$  where 'm' represents slope and 'b' represents the intercept of the y-axis. Our Y value represents Distance D and our X value represents Time<sup>2</sup> or T<sup>2</sup>. The function is defined by the equation  $D = mT^2 + b$ . So let's find the slope and intercept of the y-axis.

Move to cell **G3** and type 'Graph Parameters'. Type 'Slope' in **G5**. Activate the adjacent cell **H5**. Select **Insert, Function**, and select **Statistical**. Select **Slope** from the **Function Name** list and click **OK**. Click in the line for **Known Y's** and highlight the range with distance data. Click in the **Known X's** line and select the Time<sup>2</sup> data and select **OK**. The slope is approximately 2. Type "Intercept" in cell **G6** and move to cell **H6**. Select **Insert, Function**, and select **Statistical**. Select **Intercept** from the **Function**

**Name** list and click **OK**. Click in the line for **Known Y's** and highlight the range with distance data. Click in the **Known X's** line and select the Time<sup>2</sup> data and select **OK**. The intercept is approximately 0.

The line has the equation  $Y=2X+0$ . The relationship between distance and time in the experiment is now  $D=2T^2$ . The graph's slope of 2 is equal to half the acceleration of the ball or the ball's acceleration down the incline is  $4 \text{ cm/s}^2$ .

#### Part 4: Using the equations

The new equation can now be used to eliminate the errors that occurred with the experimental data. We can find the correct distances the cart would obtain on this incline if experimental errors were eliminated. In cell **D3**, type 'Corrected Distances'. In cell **D4**, use the formula by typing  $=2*C4$ . Copy this formula to cells **D5:D11**. The resulting values give the line of best fit and can be added to the graph.

To add the line of best fit from the data, highlight the values in Column **D** and select copy. Click on the chart area of graph 2 and click paste. A new line of best fit is placed on the graph. *Excel* can also add a line of best fit to a chart. Click a point on the chart and then right click the mouse on the point. Select **Add Trend line**, choose the **Type** tab, click **Linear**, and click **OK**. Note the overlap confirms our work.

Now for some finishing touches. Move the graph parameters down along side the second graph. Use Print Preview to view your work. Close and move and size the charts so that it all appears on one sheet. Type your group number and names in the upper right cells. Save the worksheet. Print it by selecting **File, Print**, and ensuring that **Active Sheet(s)** is activated in the print dialogue box.

Part 5: Analysis: Extrapolation and interpolation

Answer each of the following:

- Use the second graph to estimate how far the cart will travel in 3.5 seconds and in 8 seconds. (Hint: Can you change the vertical and horizontal scales?)
- How far should the cart actually roll in 3.5 seconds and in 8 seconds? (Hint: Begin in **B12** and use the copy command as required to have *Excel* find it) Explain what you did to find the values.
- How long will it take the cart to travel exactly 80 cm? How long does it take to travel 250 cm? (Hint: Use *Excel* and start in **D13**. Remember, you are solving for time T in  $D = 2T^2$ .)
- For a steeper incline, the following data was recorded for a cart rolling down it.

## ACCELERATED MOTION EXPERIMENT

Distance versus Time Data

DISTANCE(cm)	TIME(s)
0	0
2.9	1
11.5	2
25.8	3
49.1	4
73.2	5
104.5	6
149.4	7

Use the previous *Excel* worksheet to find the equation of the function again using the corrected distances. Completely update the charts and repeat each of questions 1, 2, and 3.

## Overview of Activity 6

In this activity, students will use given data and previously collected data to create scatter plots, add a line of best fit, and determine a value for the correlation coefficient.

They will then analyze these items in response to various questions.

### Objectives

#### Statistical topics.

Students will create a scatter plot and insert the line of best fit or regression line.

They investigate the equation of the regression line using the means and standard deviations of the compared data. Students will also find a value for the correlation coefficient.

#### Mathematical thinking.

Students will interpret data and analyze trends based on scatter plots, the line of best fit, and the correlation coefficient. They will conjecture potential relationships between variables. Students will also write the equation of the regression line and use it to predict other values.

#### Technology

Students will enter data, create scatter plots, and add a line of best fit to it. They will use statistical functions to find mean, standard deviation, and a value for the correlation coefficient. They will use the **cut**, **copy**, and **paste** functions to arrange data and copy formulas.

## Activity 6

Present students with the following situation:

A university professor would like to study the relationship between a student's performance on the Math Skills Inventory (MSI) administered upon entry to Memorial University and their first year math average. The following data was recorded:

MSI score (x)	70	50	60	80	75	65
First Year Math Avg.(y)	75	50	70	80	80	70

Create a scattergram and insert a trend line. Find the equation of the trend line.

Find the value of the correlation coefficient and interpret it.

#### Part 1: Scattergram with trend line

Open an *Excel* workbook and enter the above data in two columns with the appropriate headers. Activate the Chart Wizard. In step 1, select an XY (Scatter) chart of the type "Scatter. Compares pairs of values" and click **Next>**. For step 2 (Source Data), select the **Series** tab, **Add**, and enter MSI scores as **X Values** and the first year averages as the **Y Values** and click **Next>**. Step 3 allows for many formatting options at the same time. Click the **Titles** tab and enter the appropriate titles and labels. In the **Gridlines** tab, select **Major axis** for both X and Y axes. Deactivate the legend under the **Legends** tab and click **Next>**. Select **As object in** option in step 4 and **Finish**. Finally, click on each axis and then right click to format the axis. Ensure the scale goes to 100 for each.

You are now ready for a trend line. Click on a data value in the chart and then right click the mouse on this point. Select **Add Trend line**, the **Type** tab, click **Linear/Regression**, and **OK**. Note the trend line only moves between the highest and lowest data values. The trend line can also be extended to help with the interpolation or

extrapolation of data. Click on the trend line and right click the mouse. Select **Format Trend line** and click the **Options** tab. In Forecast, enter 10 in Forward, 30 in Backward, and click **OK**. Note how the line now extends.

### Part 2: About regression lines and correlation

A regression line has two properties that we can use to quickly find its equation. First, it passes through a point (ordered pair X and Y) where the x-coordinate is the mean (average) of the x-values and the y-coordinate is the mean (average) of the y-values. Second, to get the slope, divide the standard deviation of the y-coordinates by the standard deviation of the x-coordinates and the multiply by the correlation coefficient.

The correlation coefficient provides a mathematical measure of the linearity of the values but does not provide information about cause and effect. A high negative correlation approaches a value of -1 while a high positive correlation approaches a value of 1. If the value of the correlation coefficient is close to zero, then no real linear relationship exists. The equation for correlation coefficient (r) is:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y}$$

That is, the correlation coefficient for two sets of data is equal to the mean of products minus the product of the means divided by the product of the standard deviations.

### Part 3: Finding the equation of the regression line

To get the equation of the regression line, you need to find the mean, standard deviation, and correlation coefficient. Click on **A** in column **A** and insert a column. In cells A10, A11, and A12 type Mean, Standard Deviation, and Correlation Coefficient respectively. Adjust the width of the column to fit the headings by clicking the line to right of column **A**.

In **B10** select **Insert, Function, Statistical**, select **AVERAGE**, and choose **OK**. Enter the range of MSI data in the **Number 1** box and click **OK**. Copy this formula to **C10**. In **B11** select **Insert, Function, Statistical**, select **STDEVP**, and choose **OK**. Enter the range of MSI data in the **Number 1** box and click **OK**. Copy this formula to **C11**. In **B12** select **Insert, Function, Statistical**, select **CORREL**, and choose **OK**. In the **Array 1** box enter the range containing MSI data. Enter range for Math Averages in the **Array 2** box and click **OK**. The point the line passes through is (mean of X values, mean of Y values). It should be the point (70,73.333).

Type "Slope" in to cell **A13**. Again, the slope is the standard deviation of the y values divided by the standard deviation of the x values and then multiplied by the correlation coefficient. To get the slope, click cell B13 and type  $=\text{(C11/B11)}*\text{B12}$ . The slope should be 0.794118.

To write the equation of the line, recall the form  $Y - Y_1 = m(X - X_1)$ . Substitute the point above for  $(X_1, Y_1)$  and the slope for 'm'. The resulting equation for the regression line is  $Y = 0.7941X + 17.745$ . We can now find the exact value of Y by substituting any desired value for X.

#### Part 4: Data Analysis

Move to Sheet 2 of this *Excel* workbook and enter the following data in two columns with the appropriate headers:

Air Temperature C° (x)	5	3	15	10	2	0
Monthly Heating Bill (y)	165	175	90	105	180	200

Use the format of the previous exercise to create a scattergram and insert a trend or regression line. Find all the necessary data to determine the equation of the regression line. Label and type the equation of the regression line on the worksheet. Use Print

Preview to view your work. Close and move and size the charts so that it all appears on one sheet. Type your group number and names in the upper right cells. Save the worksheet. Print it by selecting **File, Print**, and ensuring that **Active Sheet** is activated in the print dialogue box. Submit this sheet with your answers to each of the following questions:

1. Is this an example of a positive or a negative correlation?
2. Study the data sets for both the situations explored. How could you estimate whether a correlation would be positive or negative from the data?
3. (a) Estimate the cost of heat for a month when the average temperature was  $8^{\circ}\text{C}$ .  
(b) What is the actual cost to the nearest whole dollar?
4. (a) Estimate the average temperature for a month that would allow the cost of heat to be zero.  
(b) What is the actual average temperature that could yield this cost?
5. In activity 2 you drew various charts and looked for a relationship between Hours of Study and Math Averages. Return to the Initial Data sheet from activity 1.  
(a) Do you feel there is a relationship between Hours of Study and Math Averages?  
(b) Create a chart and use the correlation coefficient to support or dismiss your beliefs.  
(c) Do you feel there is a relationship between Height and ability to shoot free throws? Create a chart and use the correlation coefficient to support or dismiss your beliefs.

## References

- Arganbright, D. E. (1992). Using spreadsheets in teaching statistics. In Gordon, F. & Gordon, S.(Ed.), Statistics for the Twenty-First Century, (pp. 226-242). Washington, DC: Mathematical Association of America.
- Brush, T. A. (1997). The effects of student achievement and attitudes when using integrated learning systems with cooperative pairs. Educational Technology Research and Development, 45(1), 51-64.
- Cobb, P., Boufi, A., McClain, K., & Whitenack, J. (1997). Reflective discourse and collective reflection. Journal for Research in Mathematics Education, 28(3), 258-277.
- Department of Education (1997). Atlantic Canada Mathematics Curriculum Teachers' Draft Guidelines. Government of Newfoundland and Labrador.
- Dreyfus, A., Feinstein, B., & Talmon, J. (1998). The advantages and problematics of using the electronic spreadsheet in biology teaching as perceived by actively engaged teachers. Journal of Educational Computing Research, 19(1), 67-81.
- Forte, J. A. (1995). Teaching statistics without sadistics. Journal of Social Work Education, 31(2), 204-217.
- Garfield, J. (1993). Teaching statistics using small-group cooperative learning. Journal of Statistics Education, 1(1), [on line]. Available: <http://www.amstat.org/publications/jse/v1n1/garfield.html>
- Giraud, G. (1997). Cooperative learning and statistics instruction. Journal of Statistics Education, 5(3), [on line]. Available: <http://www.amstat.org/publications/jse/v5n3/giraud.html>
- Hall, A. G. (1995). A workshop approach using spreadsheets for the teaching of statistics and probability. Computers & Education, 25(1/2), 5-12.

Hoffer, A. R. (1982). Development of teaching mathematics at school level. In Grey, D. R., Holmes, P., Barnett, V., & Constable, G. M.(Ed.). Proceedings of the First International Conference on Teaching Statistics. (pp. 2-63-268). University of Sheffield Printing: Teaching Statistics Trust.

Hogg, R. V. (1992). Towards lean and lively courses in statistics. In Gordon, F. & Gordon, S.(Ed.), Statistics for the Twenty-First Century, (pp. 3-13). Washington, DC: Mathematical Association of America.

Hunt, N. & Tyrrell, S. (1995) Discovering Important Statistical Concepts using Spreadsheets [on line]. Available: <http://www.nis.coventry.ac.uk/research/discus/discus.home.html>

Keeler, C. M. & Steinhorst, R. K. (1994). Cooperative learning in statistics. Teaching Statistics, 16(3), 81-84.

Keeler, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course. Journal of Statistics Education, 3(2), [on line]. Available: <http://www.amstat.org/publications/jse/v3n2/keeler.html>

Krulik, S., & Rudnick, J. A. (1994). Reflect...for better problem solving and reasoning. Arithmetic Teacher, 41(6), 334-338.

McManus, S. M., & Gettinger, M. (1996). Teacher and student evaluations of cooperative learning and observed interactive behaviours. The Journal of Educational Research, 90(1), 13-22.

Mitchell, M. (1997). The use of spreadsheets for constructing statistical understanding. Journal of Computers in Mathematics and Science Teaching, 16(2/3), 201-222.

Nash, J. C., & Quon, T. K. (1996). Issues in teaching statistical thinking with spreadsheets. Journal of Statistics Education, 4(1), [on line]. Available:

<http://www.amstat.org/publications/jse/v4n1/nash.html>

National Council of Teachers of Mathematics (1989). Curriculum and Evaluation Standards for School Mathematics. Reston, Va: Author.

National Council of Teachers of Mathematics (1992). Curriculum and Evaluation Standards for School Mathematics: Addenda Series - Data Analysis and Statistics.

Reston, Va: Author.

National Council of Teachers of Mathematics (1998). Principles & Standards Document (Draft), [on line]. Available:

<http://standards-e.nctm.org/1.0/normal/standards/princTec.html>

Perry, W. (1998). What software skills employers want their employees to possess? Business Education Forum, 52(3), 35-38.

Posamentier, A. S. & Stepelman, J. (1999). Teaching Secondary Mathematics: Techniques and Enrichment Units (5<sup>th</sup> ed.). New York: Prentice Hall.

Simon, M. A. (1995). Restructuring mathematics pedagogy from a constructivist perspective. Journal for Research in Mathematics Education, 26(2), 114-145.

Wheatley, G. H. (1992). The role of reflection in mathematics learning. Educational Studies in Mathematics, 23(5), 529-541.

Whicker, K. M., Bol, L., & Nunnery, J. A. (1997). Cooperative learning in the secondary mathematics class. The Journal of Educational Research, 91(1), 42-47.

Wilson, M. R., & Krapel, C. M. (1995). Exploring mean, median, and mode with a spreadsheet. Mathematics Teaching in the Middle School, 1(6), 490-495.

## APPENDIX A

### Background to *Excel*

Permission is given to reproduce the following activity sheets for use in an instructional setting provided that full credit to the author is acknowledged. The material is not to be reprinted for use in a course package or materials that are to be sold.

### Background to *Excel*

*Excel* is a spreadsheet application software for computers that is quickly replacing the pencil, paper, and calculator spreadsheets of earlier days. A spreadsheet is the document created by deliberately organizing information into a row and column format for easier calculation and analysis. Electronic spreadsheets allow manipulation of numeric data, text, calculating formulas, developing graphs, determining trends ('What-if' evaluation) and producing output such as tables, graphs, and reports. Other spreadsheet programs include *Quattro Pro* and *Lotus 1-2-3*. They all work in much the same way but *Excel* has the added advantage of being completely compatible with other applications in *Microsoft Office 97*. Charts, tables, graphs, etc created in *Excel* can be copied directly into *Microsoft Word* (word processing software) or *Microsoft PowerPoint* (slide presentation software) for professional looking reports and presentations.

To open *Excel*, click start in the lower left corner of your screen and move the pointer up to programs. A list of programs will appear to the right. Move the pointer to *Microsoft Excel* and click. At the top of the screen, you will see a series of menus called File, Edit, View, etc. You have just opened a workbook in *Excel*. A workbook, like an exercise book, has several sheets to it. Initially, you will see Sheet 1, Sheet 2, and Sheet 3 named at the bottom left of your workbook. You can create up to 256 sheets in a workbook and customize the names as desired. This and other customizing features will be addressed as required.

Under these menus File, Edit, etc., there should be two rows of symbols called "Tool bars." These items represent frequently used formatting tools. If they are not presently on your screen then click **View** and move down to **Toolbars**. In the list that

appears to the right, select **Standard**. Repeat the process to select **Formatting**. Note that a checkmark appears by the word **Standard** indicating that the toolbar has been activated. Now move the pointer to each of these symbols in the first row and let it rest. A brief message will appear to tell what each symbol does. You can play and activate other tool bars here but remember to leave the ones discussed turned on. Too many toolbars will obstruct your view of the worksheet.

Also in the **View** menu, ensure that the **Formula Bar** and **Status Bar** are highlighted with a check mark. You can turn these on and off to distinguish between them but leave them on for future reference.

Under the given tool bars, you will see the actual spreadsheet as in the diagram below. Each of the blocks you see is called a cell.

	A	B	C	D	E	F	G
1							
2							
3		Money					
4						\$400.00	

The word "**Money**" has been entered in cell **B3** and "**\$400.00**" in cell **F4**. Text or data can be entered into any cell. Each cell (block) is named by a letter and a number to allow us to place information in a specific column or row. It is similar to locating ordered pairs (x, y) or points from a function in space on the Cartesian plane. As you can observe, each column is labeled with a letter of the alphabet. Use the right arrow key or use the mouse to click on the right arrow over button in the bottom right corner of the screen. Note how the lettering of the columns proceeds: A, B, C, ..., Z, AA, AB, ... AZ, BA, BB, ..., IV for 256 columns. The rows are numbered from 1 to 8192. This potentially gives the user  $256 \times 8192 = 2,097,152$  cells.

In the previous diagram, you also observed a shaded area. When several cells are selected at a time, it is called a **range**. The range **D2** to **E3** or **D2:E3** has been shaded. That is, rows **2** and **3** in columns **D** and **E** are shaded. To select the range **A2:C3** simply place the pointer in **A2**, click and hold the mouse button and move to cell **C3**. The entire range will be highlighted. This will come in handy later for copying formulas among other things.

Press and hold the **Control** key **Ctrl** and then the **Home** key **Home**. This will put you in cell **A1**. Note in the box above column **A** you see **A1**. This box is called the **Name Box**. Use the arrow keys to move around some cells to see how the name changes.

In the activities that follow, you will experience the power of *Excel* for the analysis of statistics. Remember that you are simply organizing information into columns and rows to make it more manageable. Now learn to enjoy your new tool.

## **APPENDIX B**

### **Student Activity Sheets**

Permission is given to reproduce the following activity sheets for use in an instructional setting provided that full credit to the author is acknowledged. The material is not to be reprinted for use in a course package or materials that are to be sold.

## Student Activity 1

### Part 1: Preparing to collect and save data

Your consulting firm has been asked to gather the following class data for future analysis: name, age, grade level, hours spent studying math (week), math average, height (cm), and basketball free throw ability. Answer all questions honestly and take steps to ensure accurate measurements. Each student attempts ten baskets and you record the number of successes.

### Format the worksheet

Open *Excel* from the start menu as before. Move to cell **B1** and type the title **STUDENT PROFILE**. Notice that the words also appear in the block to the right of the **Name Box**. This box is called the **Formula Bar**. Press **Enter** or one of the arrow keys.

Now let's name the columns. In cell **A4** type **Name**. Some headings may be two words so you may like to enter the words in two rows as below in **B3** and **B4** or **C3** and **C4** for example.

	A	B	C	D	E	F	G
1		STUDENT PROFILE					
2							
3			Present	Hrs. Study	Math	Height	Baskets
4	NAME	Age	Grade	Math(wk)	Average	(cm)	(/10)
5	John Doe						

Keep the headings short but descriptive. If names are too long for the formatted cell move the pointer to the vertical line at the right of the column letter. When the plus sign changes to a pair of vertical parallel lines, double click it. The column is now formatted to the width of the longest entry. Proceed to enter the other column names. Enter the student names under **NAME** in cells **A5**, **A6**, etc.

### Saving the worksheet

When saving for the first time, you should create a folder to store all related files. Click the disk in the upper left corner. In the **'Save as'** dialogue box, click the third folder to the right to create a new folder to store your *Excel* files. In the **New Folder** dialogue box type **Excel Files** in the **Name** box and click **OK**. A new folder will appear. Click the folder so it appears in the **Save in** box. Name your new workbook by typing **Student Profile** in the **File name** box at the bottom and click **Save**. The workbook is now saved in your *Excel Files* folder as **Student Profile**. You should save your work regularly.

---

\*\*\* Activity 1 is an adaptation of an activity from an article entitled "Exploring Mean, Median, and Mode with a Spreadsheet" by Wilson and Krapel (1995).

### Viewing and printing the worksheet

Go to **File** and select **Print Preview**. There are no lines to work with and you may be missing a column. Select **Close**. To get a grid, select the range in which you want the lines beginning with **A3**. Select **Format, Cells, and Border**. Click **Outline, Inside**, and select **OK**. Preview your document and observe the changes. Save this work. Select **Print** and note your option to print the entire document or just select pages. Select **OK** to print a data collection sheet. To exit, select **File, Close**, and usually hit **Yes** if you made changes that you want to save. To close *Excel*, click the  in the upper right corner of the screen.

### Part 2: Collecting and entering data and formulas

With the data collection complete, open their *Excel* workbook and begin entering the data. When complete, skip a row and enter the statistical terms to be investigated as in column A below:

	A	B	C	D	E
29					
30	Maximum				
31	Minimum				
32	Range				
33	Mean(Avg.)				
34	Median				
35	Mode				
36					

In the cell adjacent to the word **Maximum** enter **=maxa(B5:B20)** and press **Enter**. This assumes that **B5** is the first element in the range and **B20** is the last. Your worksheet will likely be different. The equals sign, "=", tells *Excel* that a formula or predefined task will follow like **maxa**. After the task, open brackets to enter the range of cells that contain the numerical data separated by a colon and close the brackets. When you open the brackets you can quickly define the range by highlighting it with the mouse and closing the bracket at the end.

In the cell adjacent to **Minimum** enter **=mina(B5:B20)** and press **Enter**. In the cell adjacent to **Range** enter **=B30-B31** and press **Enter**. In the cell adjacent to **Mean** enter **=average(B5:B20)** and press **Enter**. In the cell adjacent to **Median** enter **=median(B5:B20)** and press **Enter**. In the cell adjacent to **Mode** enter **=mode(B5:B20)** and press **Enter**. Carefully study the numbers that appear.

### Copying formulas

You also need to calculate these values for the other columns of data. To use the copy command, begin by highlighting the range of cells that we just entered formulas into. Select the copy icon at the top of the page. A dashed border surrounds the range of cells. Highlight the entire range of cells that will hold the information and click the paste icon (next to copy). *Excel* performs the same calculations on each of the other columns.

### Customizing the sheet

Add some customization to your worksheet by double clicking the **Sheet 1** tab at the bottom of the sheet. The word **Sheet 1** will be highlighted. Simply type a new title for the page like **Initial Data** and hit Enter. Sheet 1 is now called Initial Data. Save this file replacing the previous worksheet.

### Copy and sort functions

Highlight all the cells used in the **Initial Data** sheet and click the **copy** icon. Select the **Sheet 2** tab at the bottom of the page and click **paste**. All the data is now transferred to **Sheet 2**. You may customize this sheet name also as before. Call it 'Working Data'.

At times, it may be easier to analyze the data by sorting it using some criteria. Highlight the entire range of data below the column headers. Do not include the calculations you made or the headers. Select **Data** and **Sort**. In the **Sort** dialogue box click the down arrow in the first **Sort by** section. Each of the column headers will be there. Select **Age** and **OK**. Sort again and as often as you please but remember to highlight the range so that the overall arrangement of data is preserved. In addition to the suggestions to change data, make any changes you want to help your investigation of the data. Complete the exercises that follow.

### Part 3: Analyzing the data and making conclusions

1. Based on the data, write a definition for maximum, minimum, and range.
2. What relationship exists between maximum, minimum, and mean (avg.) in each of the columns?
3. What will happen to the maximum, minimum, and mean (avg.) if one of the entries is changed to zero? What happens if one value is made extremely big?
4. Look at the sorted data values in relation to the value that represents the mode. Write a definition for mode.
5. Look at the sorted data values in relation to the value that represents the median. Write a definition for median.
6. Change a data value to zero or to an extremely large number. What effect does this have on the mode and the median? Is this consistent with your definition?
7. Experiment and change some data elements. Explain how the median is similar to but different from the mean.
8. Which single descriptive statistic - maximum, minimum, range, mean, median, or mode - best describes each of the categories?
9. If students had been asked to name their favorite sport, could either of mean, median, or mode be used to describe the resulting data? Why?

\*\*\*\* Remember to submit a copy of your Initial Data sheet with your work.

## Student Activity 2

### Part 1: Creating and editing a chart.

You are back to work at your consulting firm. Your company wants to look for a relationship between study time and math grades. Display these relationships in a separate bar graph, line graph, and XY scatter graph. Your company would like you to summarize your observations and include them with the charts in a brief written report.

#### Create the chart

In *Excel*, open the file **Student Profile**. Copy the original data to **Sheet 3**. Click the **Initial Data** sheet tab and select entire range of data. Select **copy**; click the **Sheet 3** tab at the bottom of the page and click **paste**

Highlight the range of cells to be included in the chart. To graph Hours of Study versus Math Average, highlight the entire range of data including the column titles in columns **D** and **E**. Click the **Chart Wizard** icon and follow the four-step procedure. If the Office Assistant offers help, click **NO**.

In 'step 1 of 4 - chart type', click **Column** from the list that appears and the types of column charts will appear to the right. Select the chart type 'clustered column, compares values across categories' and then **Next>**.

In step 2, you are asked for the chart source data. The range you selected is already in the blank **opposite Data range**. Ensure that **Columns** is selected in the '**Series in**' section and click **Next>**.

Step 3 allows for some customizing of the chart. Select the **Titles** tab from the above list. For the title enter 'Math Avg. versus Hours Study (wk)'. Similarly, label the x-axis 'Hours Study(wk)', the y-axis 'Math Avg.', and click **Next>**.

The final step simply asks if you desire to have the chart created on its own page or placed in the present sheet. Select '**As object in**' and **Finish**. The chart will appear on top of the data in your worksheet. Click and hold the mouse button on a white part of the chart. When the pointer changes to a four-directional arrow, you can drag the chart to another part of the worksheet. Place it directly below the data by dragging the border.

#### Edit the chart

Click the white area of the chart window and the chart Edit window is activated. You may then click on any object in the chart to activate it. When activated, the object will have handles (small black boxes) around it. You can now make any desired changes.

Click on the upper right handle of the chart and drag the border so that it rests in the upper right cell one row below the data. Next, click on the lower left handle and drag the border to make the chart six rows and six columns larger.

We can also change fonts and move text in the chart. Activate the Edit window and click the title (a frame will surround it). Select **Font Size** from the menu above and choose 14. Next click on the frame and move it to the center. This can also be done with the labels on the axis.

The scale that appeared on the axis can also be modified. Click on the vertical axis and the handles appear again. Next, click the right mouse button and select **Format axis**. Select the **Scale** tab. Turn off the check mark in **Major unit** and type 10 in the blank. Turn off the check mark in **Minor unit** and type 5 in the blank. Select **OK**.

You may also want to highlight the axis and add gridlines to help the reader interpret the results. Activate the vertical axis line and click the right mouse button. Select **Format** and the **Patterns** tab. Click **Axis, Custom**, select **Weight**, click the thick line, and select **OK**. Repeat this process for the horizontal axis. To get gridlines, click the white area of the chart window to activate it. Click the right mouse button and select **Chart Options**. Select the Gridlines tab. The horizontal axis is **Category (X) axis**. Select only the **Minor gridlines** here and in the **Value (Y) axis**. Select **OK** and close the Edit window by clicking outside the chart. **Save** and **print** your work.

Change some data values. Notice that the chart is automatically updated to show the change. When you exit, answer NO when asked if you want to save changes to the worksheet. This will preserve the original data for part 2 of the activity.

### **Part 2: Creating line graphs and scatter plots**

Follow the set of instructions from Part 1 to create a Line chart of the type 'Line with markers at each data value' and place it under the previous one. Similarly, create a XY (Scatter) chart of the type 'Scatter, compares pairs of values'. Use only the formatting you feel is necessary to clearly display the data.

Copy your charts to a new page in the workbook. Arrange the charts so that they can all be printed on the same page. The dashed line in the sheet indicates the dimensions of the paper you are working with. You must fit the charts into this space to print them on one line. Save and print your work.

### **Part 3: Interpreting and presenting your findings**

Answer the following questions in brief paragraphs and type your solutions in **Microsoft Word**. This is a word processing program that works much like *Excel* for formatting, saving, and printing a document. As part of your response to question number three, you will be asked to copy and paste a chart into your work. Save it and print it.

To copy the chart of choice, minimize **Microsoft Word** by clicking the first of three boxes in the upper right corner with a flat bar in it. Open your *Excel* file as usual and click the white part of the chart you want. Select **Copy** and a border appears around the chart. Close *Excel* and reactivate your **Microsoft Word** document by clicking on it in the status bar by the start menu at the bottom of the screen. Move to the location that you want the chart to appear and select paste. Do not worry about the appearance or formatting of your work in Microsoft Word. Finish typing your answers, then save and print your work. The questions are as follows:

1. Describe any relationships you feel may exist between success in mathematics and hours of study. Do this without looking at the data you collected.
2. In looking at the raw data alone (not the charts), describe what you observed that supported or went against your ideas of a relationship?
3. Of the three types of charts explored, which do you feel best displays and supports your ideas? Why?
4. The sample of data came from your class. Could it be used to describe all the students in the school, the province, or Canada? Support your views in each case with some logical argument.

## Student Activity 3

### **Part 1: Creating the stem-and-leaf plot**

Open your *Excel* workbook. Select **Insert** and **Worksheet** to add a new sheet to the workbook. Click the Sheet 4 tab to begin working on a new sheet and change the name of it to Stem-and-Leaf plot.

A stem-and-leaf plot allows for organizing data into classes for analysis while preserving the original data values. In this activity, we will use the math average data collected for Activity 1. The data are marks out of 100 so the stem or fixed items in the data are the tens digits. Type **Leaves** in cell **A10** and type **Stems** in **A11**. In columns **B11**, **C11**, **D11**, ... , **L11** enter 0, 1, 2, ... , 10 as below.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
:												
:												
9												
10	Leaves											
11	Stems	0	1	2	3	4	5	6	7	8	9	10
12												

The stem labels represent the intervals like 0 to 9, 10 to 19, and 20 to 29 respectively. To separate **Leaves** and **Stems** as above, select the range **A11** to **L11** for formatting. Select **Format**, **Cells**, **Border**, and double line in **Line Style**. Click only the top border and **OK**.

To record a data value, we now enter the ones digit above the appropriate tens digit. For example, if 34 were the first data value, enter 4 above the stem 3. For 30, you place a 0 above the 4 in the 3 stem. An 85 would result in a 5 over the 8. Continue to enter the data in this manner.

### **Part 2: Formatting the stem-and-leaf plot**

Starting in rows 10 and 11 should allow you to do most stem-and-leaf plots with a small data set. If you need more rows, simply go to row 1 or any place above the highest entry and select **Insert** and **Rows**.

To close up the space between the columns, move the mouse pointer to the line to the left of each column letter. When the pointer changes to a set of vertical parallel lines, click it. The column is now formatted to the width of the longest entry in the column.

It is often easier to read the chart when a vertical line separates each column. Highlight the range of cells from **B11** to **L11** and up to highest row used. Select **Format**, **Cells**, **Border**, and a solid single line in **Line Style**. In the **Border** section of this dialog box, select each of the vertical line options and **OK**.

Select **File** and **Print Preview** to view the results. Close. In cell **A1**, enter the title "Stem-and-Leaf Chart for Math Averages". Select any cell in the upper right corner of the page and enter your group number and the names of those in your group in the cells below.

### **Part 3: Drawing a histogram**

*Excel* will automatically sort data into several predefined ranges to create a histogram. To control the bin widths for the sorting of data, you must plan ahead. Let's sort the data values in the intervals (bins) 0 to 9, 10 to 19, 20 to 29, ..., 80 to 89, and 90 to 100. Return to your Initial Data sheet. Go to column **H** and enter the heading Bin Width in cell **H1**. To get the desired class widths, enter 9, 19, 29, 39, 49, 59, 69, 79, 89 below in the respective column H cells.

The histogram is not contained in the Chart Wizard. Instead, you must select **Tools, Data Analysis, Histogram**, and **OK**. In the **Histogram** dialogue box, click in the line for **Input Range**. Highlight the set of data values in the Math Average column and the range appears in the **Input Range**. You can type this in manually if you prefer. Next, click in the **Bin Range** line and highlight cells containing the range values. Before moving to the output options, ensure the Labels box is not activated with a checkmark. *Excel* will generate appropriate labels in the output table. In **Output options**, select **New Worksheet Ply** and ensure only **Chart Output** is activated by a checkmark. The histogram and output table will be created on a new sheet in the workbook. Rename this sheet as Histogram 1.

### **Part 4: Formatting the histogram**

The histogram can be customized like charts created with the Chart Wizard. Click the white area of the chart window to activate the chart **Edit** window. Drag the lower right handle to expand the histogram over 2 more columns and 6 rows.

Note the spaces between the bars. What chart does it look like? To remove these spaces, right click on any vertical bar. Select **Format Data Series, Options**, set the **Gap Width** to zero, and click **OK**. It now looks like a true histogram.

For some final touches, recall some work from the previous activity to personalize the histogram. Change the title of the chart from "Histogram" to "Histogram for Math Averages." Change the label on the horizontal axis to something like "Upper Bin Boundary." Make the horizontal and vertical axis a heavier line. Then, move the legend. Right click on Frequency, select **Format Legend, Placement**, click **Bottom**, and click **OK**. Now left click and hold to drag it to the lower left corner.

### **Part 5: Comparing and analyzing the charts**

Copy the stem-and-leaf plot to the same page as the histogram for comparison. Go to the page called Stem-and-Leaf plot. Highlight the required range and select copy. Return to the Histogram 1 sheet and select a cell below the histogram. Select paste. Print this sheet.

Answer each of the following questions:

1. (a) Explain how a stem-and-leaf plot facilitates finding the mode of a set of data.  
 (b) Explain how a stem-and-leaf plot facilitates finding the median of a set of data.  
 (c) Explain how a stem-and-leaf plot can be used to quickly add the data for finding the mean.
  
2. Of mean, median, and mode, which best represents this set of data? Explain why.

3. What does a stem-and-leaf plot show about the data that measures of central tendency do not?
4. Compare the stem-and-leaf plot you created to the histogram. What similarities do you notice and what information is lost in the histogram?
5.
  - (a) What is the total area under the individual bars that make up the histogram?
  - (b) What is the area under the individual bar for the range 70 to 79?
  - (c) What percent of the total area is contained in this interval?
  - (d) Of the total number of values, what percent fall in the range 70 to 79?
  - (e) For any interval containing values, find the area and repeat (c) and (d).
  - (f) Is there a relationship between the area of a region in the histogram and the frequency of values in the corresponding range? Explain your answer.
6.
  - (a) What intervals would you need to have five equal width bins?
  - (b) Go back to the Initial Data sheet and adjust (edit) the **Bin Width** column to obtain these bins. Create a new histogram on a new sheet as before. Name the sheet **Histogram 2** and print it.
  - (c) Describe how this compares to the previous histogram. Which offers a better picture of the data? Why?
  - (d) Predict what would happen if you used a bin width of five.

## Student Activity 4

### Part 1: Preparing to investigate standard deviation

Open a new workbook in *Excel* and set up a spreadsheet to match the format below:

	A	B	C	D	E
1		Calculating Standard Deviation			
2					
3		Data	Deviation	Squared Deviation	
4		X	X-Mean	$(X-\text{Mean})^2$	
5		10			
6		20			
7		30			
8		40			
9		50			
10	Sum				
11	Mean				Variance
12	Standard				Standard
13	Deviation				Deviation

The table above summarizes a popular set of steps that can be used to find the standard deviation. First, find the sum of the five data elements. Select cell **B10**. Next, select **Insert, Function**, and select **Statistical**. Select **SUM** from the **Function name** list and click **OK**. In the Number 1 box, enter the range **B5:B9** manually or by highlighting it with the mouse. If the resulting dialogue box covers your data, click and hold on a gray part of the box and move it as required. What would you usually do to the sum to find the mean?

For now, repeat the above process to find the mean and standard deviation in adjacent cells **B11** and **B12**. These are found in **Statistical** functions as **AVERAGE** and **STDEV** respectively. Check that the value for the mean concurs with your definition of average.

You now know the value for standard deviation without any real labor. Let's continue the process in its calculation in hopes of receiving enlightenment towards what it measures. Defining a cell or range of cells with a name is often helpful for calculations. The name can then be used in a formula and this eliminates always having to enter the range.

To name a cell like **B11** as "**mean**", click the cell (or highlight the range). Select the down arrow in the name box and type "**mean**" and press **Enter**. The changing value that appears in **B11** can now be accessed in formulas with the word **mean**.

---

\*\*\* Activity 3 is an adaptation of an activity entitled "Discovering Important Statistical Concepts using Spreadsheets" by Hunt and Tyrrell, 1995 at [http://www.mis.coventry.ac.uk/research/discus/discus\\_home.html](http://www.mis.coventry.ac.uk/research/discus/discus_home.html)

In column C, calculate the deviations from the mean. Click cell C5 and type in **=B5-mean** and **Enter**. Note the value. Copy this formula to C6:C9. Some of the values will be positive and some will be negative. Copy the formulas for **Sum** and **Mean** to their adjacent cells and note the values.

To find a more usable mean, square each of the deviations from the mean in column D. Click cell D5 and type **=(B5-mean)^2** and **Enter**. Note the value. Copy this formula to D6:D9 and copy the formulas for **Sum** and **Mean** to their adjacent cells in D10 and D11. The mean of the squared deviations is called **variance**.

Standard deviation is the square root of **variance** or square root of the mean of the squared deviations. This kind of reverses the previous squaring. Name cell D11 'var' for variance as before. In cell D12 type **=(var)^0.5**. The value should be the same as in B12. You are now ready to change values and investigate how these changes affect mean and standard deviation.

### **Part 2: Investigating standard deviation**

Change the data values as required to complete each of the following questions:

1. (a) What effect does substantially decreasing a value have on the mean and the standard deviation?  
 (b) What effect does substantially increasing a value have on the mean and the standard deviation?
2. (a) Is the sum of the deviations from the mean always zero?  
 (b) Can all the deviations from the mean be positive? Why?  
 (c) Can all but one of the deviations be positive? Give an example.  
 (d) Describe what happens if all the original integers were negative.
3. (a) Find a set of five values so that the mean is larger than the standard deviation.  
 (b) Find a set of five values so that the standard deviation is larger than the mean.
4. For a set of five numbers record the mean and standard deviation. Add ten to each number. What happens to the mean and the standard deviation? Try another number. Is there a general rule to describe the mean and the standard deviation for addition of a constant to each data element?
5. Record the mean and standard deviation for another set of five numbers. Multiply each number by ten. What happens to the mean and the standard deviation? Try another number. Is there a general rule to describe the mean and the standard deviation when each data element is multiplied by a constant?
6. Open the initial data sheet from activity 1. Copy the data to Sheet 3 and name the sheet **Standard Deviation**. Below the names in column A, label rows for **Mean** and **Standard Deviation**. In the adjacent cells, use the predefined *Excel* functions to calculate these values for each column. Use the raw data, mean, and the value for standard deviation to describe how the data are dispersed in relation to the mean for each column.

## Student Activity 5

### Part 1: Enter the data and create a chart

Present students with the following situation:

A cart was rolled down an inclined plane for a Motion-Time experiment. The data below was collected. Enter the data into a spreadsheet. Use the given steps to find the line of best fit for the data, the equation of the line of best fit, and the equation of the actual function. Use these items to extrapolate and interpolate the information indicated.

#### ACCELERATED MOTION EXPERIMENT

Distance versus Time Data

DISTANCE(cm)	TIME(s)
0	0
2.1	1
8	2
16.2	3
35	4
42.4	5
78.3	6
94.6	7

Begin entering data in cell **A1** and do not skip any rows. Your final data entries should be in cells **A11** and **B11**. Activate the **Chart Wizard**, select the **XY (Scatter) chart** using the “Scatter with data points connected by smoothed lines” type, and click **Next>**. For **Data Range**, highlight the data in cells **A4** to **A11** and click **Next>**. Click the **Chart Title** and type “ACCELERATED MOTION”. Click and enter **Time(s)** and **Distance(cm)** in **Value(X)** and **Value (Y)** respectively. Hit **Next>**. Select **As object in** and click **Finish**. Note where the values begin on the x-axis.

We now need to input the x-values. Click on the graph to activate the edit window and click on any data value (it will change color). Right click the mouse on a data value, select **Source Data**, and the **Series** tab. Now click in the X-Values entry box, highlight (enter) the range **B4:B11**, and click **OK**. The true data values are reflected on the graph.

To make the graph more readable, we will do some editing as in previous activities. First, change the scale on the x-axis so that **Auto Maximum** is 7 and the **Major Unit** is 1. Highlight both of the axes for emphasis. To add gridlines to the graph, begin by clicking on the graph background to activate it. Right click the mouse, select **Chart Options**, and the **Gridlines** tab. Select **Major Gridlines** for both the X and Y axes and click **OK**. Size and move the graph so that the upper right corner is in cell **A15** to prepare to analyze the data.

### Part 2: Analyze the data and redo the chart

When analyzing experimental data, we would like to have a straight-line graph. The graph of the data appears to be a segment of a parabola. This indicates that the x-axis data should be raised to a power. Square the time in column **C** to try to get a straight line.

Enter the header "**Time<sup>2</sup>**" in column **C3**. Type **Time2** in the cell. Highlight the 2, select **Format, Cells, Font**, click **Superscript**, and **OK**. In cell **C4**, type **=B4^2**. This formula may then be copied to the range **C5:C11** as in a previous activity. Another quick way to copy this is to first select cell **C4**. Then, click on the small square box in the bottom right corner of the cell and drag it down to cell **C11**. You are now ready to draw a new graph for Distance versus Time Squared.

Begin by copying the old chart. Click the chart, select **Copy**, and move to cell **A35** and select **Paste**. A duplicate chart will appear for editing.

To enter the time squared data, click in the chart to activate the edit window. Activate the data points and right click a point. Select **Source Data** and click in the **X Values** line. Delete the contents of this line. Enter (highlight) the range **C4:C11** and click **OK**. We now have a straight line, but the scale on the x-axis only goes to 7.

For some final touches, edit the x-axis to change **Maximum** to 50 and the **Major Unit** to 10. Then, we need to change Time to Time<sup>2</sup> in the title and on the x-axis. Double click the cursor after the 'e' in time so it appears as a flashing vertical bar. Type a '2' and then highlight it. Select **Format, Selected Axis Title, Superscript**, and **OK**. Repeat this in the title and click outside the chart to turn off the edit box. Save the work.

### Part 3: Getting the equations

To get the equation of the function for graph one, we must first determine the equation of the line in the Distance versus Time<sup>2</sup> graph. That is, we must get the slope and y-intercept (parameters) for the equation of the line.

The equation of a line is traditionally written as  $Y=mX+b$  where 'm' represents slope and 'b' represents the intercept of the y-axis. Our Y value represents Distance D and our X value represents Time<sup>2</sup> or T<sup>2</sup>. The function is defined by the equation  $D = mT^2 + b$ . So let's find the slope and intercept of the y-axis.

Move to cell **G3** and type "Graph Parameters". Type "Slope" in **G5**. Activate the adjacent cell **H5**. Select **Insert, Function**, and select **Statistical**. Select **Slope** from the **Function Name** list and click **OK**. Click in the line for **Known Y's** and highlight the range with distance data. Click in the **Known X's** line and select the Time<sup>2</sup> data and select **OK**. The slope is approximately 2. Type "Intercept" in cell **G6** and move to cell **H6**. Select **Insert, Function**, and select **Statistical**. Select **Intercept** from the **Function Name** list and click **OK**. Click in the line for **Known Y's** and highlight the range with distance data. Click in the **Known X's** line and select the Time<sup>2</sup> data and select **OK**. The intercept is approximately 0.

The line has the equation  $Y=2X+0$ . The relationship between distance and time in the experiment is now  $D = 2T^2$ . The graphs slope of 2 is equal to half the acceleration of the ball or the ball's acceleration down the incline is 4 cm/s<sup>2</sup>.

### Part 4: Using the equations

The new equation can now be used to eliminate the errors that occurred with the experimental data. We can find the correct distances the cart would obtain on this incline if experimental errors were eliminated. In cell **D3**, type "Corrected Distances". In cell **D4**, use the formula by typing  $=2*C4$ . Copy this formula to cells **D5:D11**. The resulting values give the line of best fit and can be added to the graph.

To add the line of best fit from the data, highlight the values in Column **D** and select copy. Click on the chart area of graph 2 and click paste. A new line of best fit is placed on the graph. *Excel* can also add a line of best fit to a chart. Click a point on the chart and then right click the mouse on the point. Select **Add Trend line**, choose the **Type** tab, click **Linear**, and click **OK**. Note the overlap confirms our work.

Now for some finishing touches. Move the graph parameters down by the second graph. Use Print Preview to view your work. Close and move and size the charts so that it all appears on one sheet. Type your group number and names in the upper right cells. Save the worksheet. Print it by selecting **File, Print**, and ensuring that **Active Sheet(s)** is activated in the print dialogue box.

### **Part 5: Analysis: Extrapolation and interpolation**

Answer each of the following:

- Use the second graph to estimate how far the cart will travel in 3.5 seconds and in 8 seconds. (Hint: Can you change the vertical and horizontal scales?)
- How far should the cart actually roll in 3.5 seconds and in 8 seconds? (Hint: Begin in **B12** and use the copy command as required to have *Excel* find it) Explain what you did to find the values.
- How long will it take the cart to travel exactly 80 cm? How long does it take to travel 250 cm? (Hint: Use *Excel* and start in **D13**. Remember, you are solving for  $T$  in  $D = 2T^2$ .)
- For a steeper incline, the following data was recorded for a cart rolling down it.

#### **ACCELERATED MOTION EXPERIMENT**

##### **Distance versus Time Data**

DISTANCE(cm)	TIME(s)
0	0
2.9	1
11.5	2
25.8	3
49.1	4
73.2	5
104.5	6
149.4	7

Use the previous *Excel* worksheet to find the equation of the function again using the corrected distances. Completely update the charts and repeat questions 1, 2, and 3.

## Student Activity 6

A university professor would like to study the relationship between a student's performance on the Math Skills Inventory (MSI) administered upon entry to Memorial University and their first year math average. The following data was recorded:

MSI score (x)	70	50	60	80	75	65
First Year Math Avg.(y)	75	50	70	80	80	70

Create a scattergram and insert a trend line. Find the equation of the trend line. Find the value of the correlation coefficient and interpret it.

### Part 1: Scattergram with trend line

Open an *Excel* workbook and enter the above data in two columns with the appropriate headers. Activate the Chart Wizard. In step 1, select an XY (Scatter) chart of the type "Scatter. Compares pairs of values" and click **Next>**. For step 2 (Source Data), select the **Series** tab, **Add**, and enter MSI scores as **X Values** and the first year averages as the **Y Values** and click **Next>**. Step 3 allows for many formatting options at the same time. Click the **Titles** tab and enter the appropriate titles and labels. In the **Gridlines** tab, select **Major axis** for both X and Y axes. Deactivate the **legend** under the **Legends** tab and click **Next>**. Select **As object in** option in step 4 and **Finish**. Finally, click on each axis and then right click to format the axis. Ensure the scale goes to 100 for each.

You are now ready for a trend line. Click on a data value in the chart and then right click the mouse on this point. Select **Add Trend line**, the **Type** tab, click **Linear/Regression**, and **OK**. Note the trend line only moves between the highest and lowest data values. The trend line can also be extended to help with the interpolation or extrapolation of data. Click on the trend line and right click the mouse. Select **Format Trend line** and click the **Options** tab. In **Forecast**, enter 10 in **Forward**, 30 in **Backward**, and click **OK**. Note how the line now extends.

### Part 2: About regression lines and correlation

A regression line has two properties that we can use to quickly find its equation. First, it passes through a point (ordered pair X and Y) where the x-coordinate is the mean (average) of the x-values and the y-coordinate is the mean (average) of the y-values. Second, to get the slope, divide the standard deviation of the y-coordinates by the standard deviation of the x-coordinates and then multiply by the correlation coefficient.

The correlation coefficient provides a mathematical measure of the linearity of the values but does not provide information about cause and effect. A high negative correlation approaches a value of -1 while a high positive correlation approaches a value of 1. If the value of the correlation coefficient is close to zero, then no real linear relationship exists. The equation for correlation coefficient (r) is:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y}$$

That is, the correlation coefficient for two sets of data is equal to the mean of products minus the product of the means divided by the product of the standard deviations.

### **Part 3: Finding the equation of the regression line**

To get the equation of the regression line, you need to find the mean, standard deviation, and correlation coefficient. Click on **A** in column **A** and insert a column. In cells A10, A11, and A12 type Mean, Standard Deviation, and Correlation Coefficient respectively. Adjust the width of the column to fit the headings by clicking the line to right of column A.

In **B10** select **Insert, Function, Statistical**, select **AVERAGE**, and choose **OK**. Enter the range of MSI data in the **Number 1** box and click **OK**. Copy this formula to **C10**. In **B11** select **Insert, Function, Statistical**, select **STDEV**, and choose **OK**. Enter the range of MSI data in the **Number 1** box and click **OK**. Copy this formula to **C11**. In **B12** select **Insert, Function, Statistical**, select **CORREL**, and choose **OK**. In the **Array 1** box enter the range containing MSI data. Enter range for Math Averages in the **Array 2** box and click **OK**. The point the line passes through is (mean of X values, mean of Y values). It should be the point (70,73.333).

Type "Slope" in to cell **A13**. Again, the slope is the standard deviation of the y values divided by the standard deviation of the x values and then multiplied by the correlation coefficient. To get the slope, click cell **B13** and type  $=\text{(C11/B11)}*\text{B12}$ . The slope should be 0.794118.

To write the equation of the line, recall the form  $Y - Y_1 = m(X - X_1)$ . Substitute the point above for  $(X_1, Y_1)$  and the slope for 'm'. The resulting equation for the regression line is  $Y = 0.7941X + 17.745$ . We can now find the exact value of Y by substituting any desired value for X.

### **Part 4: Data analysis**

Move to Sheet 2 of this *Excel* workbook and enter the following data in two columns with the appropriate headers:

Air Temperature $^{\circ}\text{C}$ (x)	5	3	15	10	2	0
Monthly Heating Bill (y)	165	175	90	105	180	200

Use the format of the previous exercise to create a scattergram and insert a trend or regression line. Find all the necessary data to determine the equation of the regression line. Label and type the equation of the regression line on the worksheet. Use Print Preview to view your work. Close and move and size the charts so that it all appears on one sheet. Type your group number and names in the upper right cells. Save the worksheet. Print it by selecting **File, Print**, and ensuring that **Active Sheet(s)** is activated in the print dialogue box. Submit this sheet with your answers to the following questions:

1. Is this an example of a positive or a negative correlation?
2. Study the data sets for both the situations explored. How could you estimate whether a correlation would be positive or negative from the data?
3. (a) Estimate the cost of heat for a month when the average temperature was  $8^{\circ}\text{C}$ .  
(b) What is the actual cost to the nearest whole dollar?

4. (a) Estimate the average temperature for a month that would allow the cost of heat to be zero.  
(b) What is the actual average temperature that could yield this cost?
  
5. In activity 2 you drew various charts and looked for a relationship between Hours of Study and Math Averages. Return to the Initial Data sheet from activity 1.
  - (a) Do you feel there is a relationship between Hours of Study and Math Averages?
  - (b) Create a chart and use the correlation coefficient to support or dismiss your beliefs.
  - (c) Do you feel there is a relationship between Height and ability to shoot free throws? Create a chart and use the correlation coefficient to support or dismiss your beliefs.







