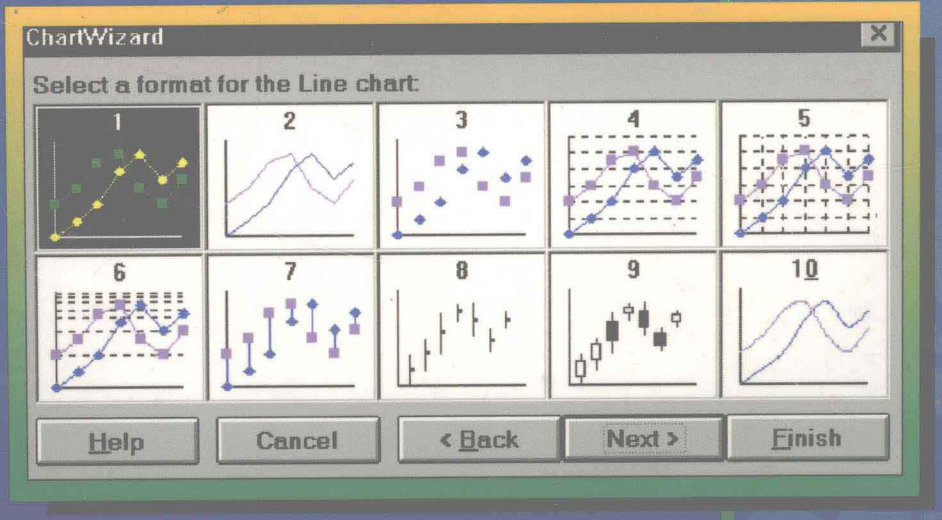
Learning Business Statistics *with* Microsoft® Excel



John L. Neufeld

Learning Business Statistics *with* Microsoft® Excel

John L. Neufeld

University of North Carolina at Greensboro

PRENTICE HALL
Upper Saddle River, New Jersey 07458

Library of Congress Cataloging in Publication Data
Neufeld, John L.
Learning Business Statistics with Microsoft Excel / John L. Neufeld.
p. cm.
Includes Bibliographical reference and index.
ISBN 0-13-234097-6
1. Commercial Statistics. 2. Microsoft Excel for Windows (Computer File) 3.Commericial Statistics--Computer Programs. I. Title.
HF1017.N47 1997 96-4136
519.5'0285'5369--dc20 CIP

Director of Production and Manufacturing: Joanne Jay
Managing Editor: Katherine Evancie
Project Manager: Susan Rifkin
Manufacturing Manager: Vincent Scelta
Manufacturing Buyer: Alana Zdinak
Design Director: Patricia Wosczyk
Interior Design: Mediamark
Editor-in-Chief: Richard Wohl
Aquisitions Editor: Tom Tucker
Executive Marketing Manager: Nancy Evans
Cover Design: Bruce Kenselaar
Production Coordinator: David Cotugno

Printed in the United States of America
10 9 8 7 6 5 4 3 2

ISBN 0-13-234097-6

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall of Canada, Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice Hall do Brasil, Ltda., *Rio de Janeiro*

Learning Business Statistics *with* Microsoft® Excel

To Karol, Paul, Adam, and Margot. Thanks.

Preface

▼ To the Student

This book is intended for the student learning business statistics who likes the idea of using Excel to help that learning. If you already know Excel, using that knowledge to help you learn statistics probably seems like a good idea. If you don't know Excel, or don't feel fully comfortable with it, using this book will enable you to learn it and become better at using it. No knowledge of Excel is assumed and the use of Excel is explained with clear and easy to follow steps.

There is a *synergy* between knowing computers and knowing statistics. Knowing how to use computers not only will help you solve practical statistical problems, *it can also help you understand statistical concepts.* The computer can empower you. Like any good tool, a computer enables you to do things you could not do without it. When you learn statistics, a computer will enable you to literally see statistical concepts which otherwise could only be imagined. An example of this is a sampling distribution, a concept involving the selection of all possible samples from a population. This fundamental concept of statistics would be much easier to understand if you didn't have to depend entirely on your imagination. In a lesson in this book you will be guided in using Excel to select 1,000 samples from a population and in a single graph observe how a sampling distribution actually looks. You will then easily explore how changes in the population affect this sampling distribution. Concepts which might otherwise seem abstract can be brought to life through the use of computers and thereby made easier to understand. Too many people approach statistics as a topic to memorize rather than to understand, and these people are never comfortable with the use of statistics. A good working knowledge of statistics doesn't require a rigorous mathematical approach, but it does require an *intuitive* understanding. Using computers to explore statistical concepts can make that intuition easier to acquire.

There are other good reasons why a knowledge of computers in general, and Excel in particular, would be useful. Too often, people try to learn a program like Excel by itself without a place to apply it. This is a very inefficient way to learn how to use a computer. For most people, computers

are not so inherently interesting as to be an appropriate subject in themselves. You don't learn how to hammer by studying hammers. You learn how to hammer by needing to hit a nail. You learn computers by having a need the computer can help you fulfill. Do you need to learn business statistics? That's the nail. That's what you should focus on. Using this book will enable you to learn Excel not for its own sake, but because it's going to help you learn business statistics.

If you already have a good working knowledge of Excel, you may feel that some of the steps in the beginning are too detailed. This feeling will not persist. My assumption is that by the end of the book, everyone will have learned Excel. As the book progresses, the Excel instructions become less detailed and less effort is made to explain the features of Excel. Excel is rich with features, and there are many alternative ways of accomplishing the same end. Even if you feel you know Excel well, you may well learn some new tricks from this book which will help you use Excel more efficiently.

- This is not a book you just sit down and read at your desk (unless your desk has a computer). *This is a book you need to use with a computer.* Many of the paragraphs of this book have bullets (like this one). These paragraphs contain step-by-step instructions for you to carry out on your computer.

Using this book will help you learn business statistics, but it won't make the process automatic. Although the approach is often step-by-step, you must not follow the steps mindlessly. The objective of the steps is not to get the computer to display a particular number, it's to help you understand either a statistical concept or the process of solving a statistical problem. After you have completed a series of steps, reflect on what you have done. Read carefully the paragraphs which do not have bullets; they are designed to help you integrate your computer experiences with your growing understanding of statistical concepts and methods. Do the exercises at the end of each chapter *before* you look up the answers. Applying the methods you have learned in the text to a new problem can sometimes seem like a struggle, but it is in this struggle that learning occurs. You can test your understanding by doing the problems at the end of each chapter after the first. Getting through a chapter and yet being unable to do the problems indicates that you need to devote more thought to what the chapters are leading you through. Answers to most of the questions are in an appendix and are provided to enable you to check your work and thus your level of understanding.

This book is primarily intended for use in an undergraduate or MBA introductory college business or economics statistics course. You may find this book's approach helpful for such a course even if it is not required

by the instructor. It also can be used by those who wish to learn the fundamentals of business statistics (and Excel) and who are not taking a formal course.

▼ To the Professor

I have had almost twenty years of experience as both a teacher and a user of statistics. During that time there has been tremendous advance in computing technology. This has brought about enormous change in the practice of statistics in business and by researchers. Sadly, it has had almost no effect on the pedagogy of statistics. As every college instructor knows, this is not owing to a lack of commercially available texts. You, like I, are probably inundated with textbooks and advertisements for textbooks, most of which are almost identical in their pedagogical approach. The problem is not that computers are ignored. Hardly a text today is published without an optional or accompanying data diskette. Most texts illustrate computer printouts in their pages. Many have supplements dealing with a particular software package—Minitab, SAS, and, increasingly, Excel. All of this, however, is very much of an add-on. The pedagogy of most texts has hardly changed over the past 20 years, and doesn't really depend on computers at all.

Most statistical programs have really been developed with the researcher or statistical user in mind, not the student. This is enormously valuable to those using statistics because it makes it easy to do complex statistical analysis on larger and larger sets of data. The intended users of these packages are expected to already understand the statistical concepts implemented in their code. It is, of course, important for students to gain some familiarity with these tools since that is how statistics is actually done. Unfortunately, these programs may hinder learning for the very reason that they are so easy to use. Why struggle with an understanding of regression when it can be done with a few mouse clicks? The response to this has tended to be a separation between *learning* statistics, which is done by hand or with a calculator, and *using* statistics, which is done with a computer. Even the widespread use of statistical tools has had remarkably little effect on pedagogy. How many textbooks continue to use the critical value approach as the primary method of teaching hypothesis testing even though the standard for computer programs is p values? Although both approaches are obviously equivalent, p values are more intuitive, but they are not as convenient to use with statistical tables. Statistical tables have more and more become used only by students. For most uses of statistics (certainly not all), they are an anachronism foisted upon students.

How is the approach used in this book different? Excel differs from most special purpose statistical packages in its versatility in handling all types of numeric manipulations, and it is easier for most computer novices to use than a statistical program. Consider the generation of a new variable. To do this in most statistical programs, you merely write an equation showing the new variable as a calculation of existing variables. Those of us who are used to using statistics have no difficulty visualizing this as operating on each of many observations generating a new value for each. Compare this to the approach with Excel. Each observation is seen as a row in a spreadsheet. A formula is entered into a new cell to calculate a new value using the values in existing cells. This value appears as soon as the formula is finished. That formula is then copied down the column, and the new value is immediately visible. This is inherently more visual and more intuitive than the approach taken by most statistical packages. Although it may not be more convenient for the researcher, it is preferable for the student.

Excel contains procedures much like those in the standard packages which automatically do such analyses as regression. These are fully covered in this book. What sets the approach of this book apart, however, are the other ways in which Excel is used to help students understand statistical concepts. Consider Chapter 8. Students explore the concepts of Type I and Type II error by performing 1,000 hypothesis tests in which the Null Hypothesis is either known to be correct or known to be wrong. A table showing the outcome of these tests is constructed, and the student actually sees the impact of significance level on incidence of Type I error and the impact of the difference between hypothesized and actual population parameter on the incidence of Type II error. Or consider Chapter 14 . If you, like many of us, think it is a good idea for students to have some experience estimating a simple regression by hand, you may like the approach taken there. Rather than immediately introducing the student to the use of the Regression Tool, the computer is first used to explore regression's concepts. Students experiment with trying to fit a line to data by guessing the regression coefficient. The regression coefficients are introduced by having the student construct the type of table which would be used in an estimation by hand. The relationship between the values in that table and scatter plots are also shown. Is this better than simply calculating the coefficients by hand? Yes. First the arithmetical drudgery is eliminated, and it becomes a simple matter to change the data estimated. Students can see the difference in the calculations between data which are well-fit by a regression line and for which there is no clear relationship.

This book should be used by instructors who are interested in pedagogical innovation. It probably is best used as a supplemental text to a more

conventional textbook. Its use requires a commitment to computers, and students will have to spend many hours working through the instructions. This is probably best done in a lab setting where all students are working together, but it can be done out of class. You will have to provide reinforcement in your lectures to the concepts students are working through. The careful step-by-step approach used in this book makes it possible for students to proceed on their own, but some will have a tendency to focus too much on the steps and may not fully grasp the concepts unless you reinforce them by referring to the process they go through on the computer.

The traditional approach to teaching statistics has been exhaustively worked out and is backed by enormous quantities of instructional material. Such a course is difficult for students to learn, but it is not a particularly difficult course to teach. This is probably a factor in the subject's pedagogical inertia. The approach used in this book is new. It will undoubtedly be improved. I believe that by using it you will provide greater value to your students. Your course may become more exciting and more effective, but it will not be easier for you. It is my fervent hope that the approach taken here will be taken up by others and that we will have a choice among high quality material which makes effective use of computers in the pedagogy of statistics.

▼ Which Version of Excel?

This book was primarily written using Excel 5.0 for Windows 3.1. This is almost identical to Excel 5.0 for the Macintosh. The two versions share the same software manual. The differences arise primarily from the different mouse and keyboard used by the two systems. I have tried in the text to accommodate Macintosh users, but I have not tested this book on a Macintosh. I anticipate no problems in using it on a Macintosh, and I would appreciate any observations or comments from Macintosh users on improvements I could make for them.

Microsoft also has released Excel 7.0 for Windows 95. (There was no Excel 6.0.) Although the new version has a few minor enhancements, including cosmetics, version 7.0 is primarily the same as version 5.0. The biggest changes are in the help system and in the dialog boxes used for reading or writing files. Although most of the screen captures in this book come from version 5.0, users of version 7.0 will have no difficulty. When the differences were substantial, I have separate instructions for the two versions. Since students are more likely to know which version of Windows they are using than which version of Excel, I have referred to 7.0 as the Windows 95 version and to 5.0 as the Windows 3.1 version.

It is possible to use version 5.0 in Windows 95, however, and the windows in that case will be something of a hybrid. Since the discussions of the two separate versions are brief and together, I do not expect this to cause any problem.

▼ The Data Diskette

A data diskette is required for several of the lessons in this book. By *not* including the diskette with the book, Prentice Hall is able to charge a lower price than would otherwise be necessary. Copies of the diskette can be provided by Prentice Hall representatives. The data can also be downloaded from the internet in the form of a program in a file named *neufeld.exe*. Download the file, have a blank formatted diskette ready, and run the program from a DOS prompt or by choosing "Run" from the "File" menu of the Windows 3.1 Program Manager. The program will give you instructions and then write all of the data files to a diskette. If this program is placed on a network, students can run it individually to create their own data diskettes. The file can be downloaded either from Prentice Hall's anonymous ftp site, ftp.prenhall.com, in the directory pub/be/decision_science.d-012/neufeld/bus.stat, or from my WWW site given at the end of this preface.

▼ A Note on Student Computer Files

All of the lessons have students develop material in Excel workbooks. The instructions tell them to save their work frequently so that a computer problem will not cause them to lose what they have already done. It is best if these files are not saved to diskette while doing a lesson. It takes significantly longer to save a file to a diskette than a hard disk or network disk. When the student is finished working, the file can be copied to a diskette in situations where students cannot permanently store files on a local hard disk or network drive.

In some cases (including chapters 6 and 8) the files required to store the workbooks will become too large to fit on a diskette. Either hard disks or network drives will have to be used in these cases. Compression/decompression utilities (like PKZip, WinZip, and others) can be used to shrink the file so that it will fit on a diskette. A set of utilities to do this is included on the data diskette, and instructions for their use can be found in Appendix D. Your computing center may have alternative programs which are easier to use.

▼ Acknowledgments

Many people have helped me in developing this book by reading and working through chapters and making comments to me and by talking with me about computers, statistics, and pedagogy. At my own institution I owe a particular debt to Elizabeth Anthony, Andy Brod, Mike Evans, Terry Seaks, and Judy Tuttle. I am also grateful for the reviews and comments I have received from those at other institutions, including Kent S. Borowick at Baylor, Teresa Dalton at University of Denver, Paul J. Fields at the Naval Postgraduate School, Phillip C. Fry at Boise State University, Mehran Hojati at University of Saskatchewan, and Joseph Zaremba at SUNY Geneseo. Thanks also to Tom Tucker, the Decision Sciences Editor at Prentice Hall, who has been patient and encouraging, and to Diane Peirano, Audrey Regan, and Patricia Wosczyk at Prentice Hall.

My greatest thanks go to the hundreds of students who have shared classrooms with me and from whom I have learned more than I have ever taught.

I wish I could promise that this book is without error. I know from experience the frustration to student and professor alike of encountering errors in textbooks, especially statistics texts. It's only with this project that I have learned how truly difficult it is to eliminate all errors in a book of this sort. It is my intention to have a list of errata available at my World Wide Web site, whose address is given below. Please check it if you think you have found an error, and let me know if you have discovered a new error.

If you have any comments, suggestions, or criticisms you would like to share with me, I will welcome them.

John L. Neufeld
Department of Economics
University of North Carolina at Greensboro
Greensboro, NC 27412

email: **john_neufeld@uncg.edu**

WWW: **http://www.uncg.edu/bus_stat/**

Brief Contents

Detailed Contents