Statistical Thinking in Business

Second Edition o in Busines: thinking in Business Thinking in Business cal Thinking in Business cal Thinking in Business stical Thinking in Busines tistical Thinking in Bus ness hti tical Thinking in Business atistical Thinking in Business Statistical Thinking in Business Statistical Thinking in Business Statistical Thinking in J Statistical Thinking Statistical Thinking Statistical Thinking Statistical Thinking Statistical Thinking

istical Thinki

J. A. John D. Whitaker D. G. Johnson



Statistical Thinking in Business

Second Edition

J. A. John

University of Waikato New Zealand

D. Whitaker

University of Waikato New Zealand

D. G. Johnson

Loughborough University



Published in 2006 by Chapman & Hall/CRC Taylor & Francis Group 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL 33487-2742

© 2006 by Taylor & Francis Group, LLC Chapman & Hall/CRC is an imprint of Taylor & Francis Group

No claim to original U.S. Government works Printed in the United States of America on acid-free paper 10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 1-58488-495-9 (Softcover) International Standard Book Number-13: 978-1-58488-495-8 (Softcover) Library of Congress Card Number 2005005695

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

John, J. A.

Statistical thinking in business / J.A. John, D. Whitaker, D.G. Johnson. -- 2nd ed.

n. cm.

Rev. ed. of: Statistical thinking for managers. c2001.

Includes bibliographical references and index.

ISBN-13: 978-1-58488-495-8 (hardcover : acid-free paper)

ISBN-10: 1-58488-495-9 (hardcover : acid-free paper)

1. Mathematical statistics. I. Whitaker, D. (David) II. Johnson, D. G. (David G.), 1943- III. John, J. A. Statistical thinking for managers. IV. Title.

QA276.12.J375 2005 001.4'22--dc22

2005005695



Visit the Taylor & Francis Web site at http://www.taylorandfrancis.com and the CRC Press Web site at http://www.crcpress.com

PREFACE

Many managers, whether in manufacturing or processing industries, in service organisations, government, education, or health, think that statistics is not relevant to their jobs. Perhaps this is because they view statistics as a collection of complicated techniques for which they see little application.

However, all business activities are subject to variability, and hence uncertainty, and managers have to make decisions in this environment. Managers need to understand the nature of variability. Otherwise they end up making wrong or inappropriate decisions that can be very costly to their organisations. As well as understanding variability, managers should understand that good decision making involves the use of meaningful information. Without good information, or data, managers have to resort to gut feelings or hunches, neither of which can be relied on. Statistics tells us how to deal with variability and how to collect and use data so that we can make informed decisions. All managers need to be able to *think statistically* and to appreciate basic *statistical ideas*.

The aim of the book is to introduce you to this way of thinking and to some of these ideas. Statistical thinking should be an integral and important part of any manager's knowledge. H. G. Wells' said that statistical thinking would one day be as necessary for efficient citizenship as the ability to read and write. It should not be regarded as some separate activity done rarely or left to others to do. Most managers would accept that they need a variety of different skills to inform and enhance their decision making. For example, being able to correctly assess the financial implications of a certain decision and being able to see how it might affect staff and colleagues are undeniably important. But these are, in a sense, secondary issues. If you have failed to find the right course of action in the first place by not being able to fully appreciate the significance of the available data, then knowing what this action will cost to implement is of little consequence.

^{*} Mankind in the Making, Chapman and Hall: London, 1914, p. 204. This is not exactly what Wells wrote, but it is commonly paraphrased in this form.

xii Preface

Unlike many books concerned with statistics for business or managers, this book does not emphasise mathematics or computation. Statisticians use mathematics extensively in their work, and the computer has become an indispensable tool for the analysis of data. However, our purpose is not to set out the mathematical basis of statistics or to give a series of statistical techniques that require considerable computation. Statistical ideas can be understood with the minimum of arithmetical work, and this is what we aim to do in the book. This does not necessarily make the book simple or basic. Recognising how statistics and statistical thinking fit into the overall business picture is important, and it requires careful thought and understanding. For example, being able to calculate a standard deviation or the equation of a line of best fit is not an end in itself. It is just a tool for describing data and is the beginning of a process of understanding and diagnosis, leading to effective decision making.

A thorough statistical analysis can involve complex ideas and extensive computation, for which specialised knowledge and computer programmes are indispensable. At times, therefore, additional assistance from a qualified statistician will be needed, but this need has to be recognised. Knowing what you can confidently do for yourself and where expert advice is needed are key characteristics of effective management.

It will be necessary to draw graphs, construct tables, and do some calculations. However, all these things can be done quite easily on the computer. In the book we make extensive use of the spreadsheet package Microsoft Excel. Spreadsheets are now standard on all computers and are, therefore, readily accessible. They are also widely used in business, and not just for statistical work. If you have not used a spreadsheet package before or are new to Excel, you may find the introduction given on the CD-ROM a useful starting point. In the book you will be shown how to use Excel for analysing and plotting data. It will be well worthwhile working though the examples and exercises involving Excel on a computer. Excel is easy to use; the more familiar you become with it, the easier it will be to work though those sections involving its use.

A few of the graphs and calculations we use in the book cannot be easily done with Excel 2003. In particular, there is no easy way to obtain a stem and leaf plot or a box plot. We have, therefore, provided an add-in that allows you to construct these plots and has a couple of other features. A copy of this *STiBstat* add-in is on the CD-ROM. We suggest you load it into Excel following the instructions given in the introduction to Excel on the CD-ROM.

One of the innovative features of this book is the inclusion of a number of hands-on exercises and experiments. It is another way we hope to encourage a process of learning by involvement. The questions contained within each chapter are there to actively involve you in the learning process by making you think about important issues to check your understanding. The experiments are a part of this same process, providing (we hope!) an active demonstration of a key concept through participation in a meaningful simulation. The most important of these experiments is the *red beads experiment*, which is introduced in Chapter 1 and forms the basis of many of the important ideas introduced in the book. To get the most out of this experiment, it should be run in the classroom so that you can see actual results occurring. If this is not possible, we have

Preface xiii

provided a spreadsheet version of the experiment that you can run in Excel. A copy of this spreadsheet is on the CD-ROM. Instructions for using the spreadsheet are provided on the opening sheet. If you cannot participate in a classroom experiment, then substitute your own results from the beads spreadsheet.

In later chapters, you will find four other experiments that can be run either in the classroom or, in most cases, on your own. Two of these, the *dice experiment* and the *quincunx experiment*, have spreadsheet alternatives you can get from the CD-ROM and run so that you can simulate the actual experiment for yourself. Again, you can use your own results instead of those produced in the classroom or contained in the book. Each experiment demonstrates a key learning point. Think carefully about what these learning points are, and try to relate them to real situations that you have experienced.

In addition to the experiments, we provide a number of demonstrations of important statistical techniques we find useful in our teaching. These are explained in the appropriate places of the text. They also are available on the CD-ROM.

In the text you will find numerous questions to answer. They are there to encourage you to think about the various issues involved. You will often get a greater understanding by first thinking things through for yourself. We strongly advise that you attempt to answer these questions as you work through each chapter. Suggested answers will be found on the CD-ROM. At the end of each chapter are exercises for you to do. You will also find over 300 multiple choice questions on the CD-ROM, designed to test your understanding of the statistical issues involved. Some of them are relatively simple, but many will require some thought. A few will require simple calculations. You can check your answers either as you do each question or after you have done a number of them.

Most of the examples in the book are based on our experiences working with a range of organisations throughout New Zealand and the United Kingdom or on those of our colleagues and associates. For confidentiality reasons we do not name those organisations, but we have tried to preserve the realism of the examples. Although the context often relates to New Zealand, most, if not all, examples can be put into the context of any other country. We encourage you to think about similar problems or examples in your country.

The material in the book can be used in a number of different ways. All of it can be adequately covered in a one-semester course of about 36–40 hours. Additional tutorial time should also be allowed in which a selection from the exercises at the end of each chapter can be used. We also suggest you attempt some of the multiple choice questions on the CD-ROM. We see no need for separate computer laboratory sessions, for our focus throughout is on developing your understanding of the role statistics plays in business, rather than on the arithmetic detail. The book can be used for a shorter course by skipping through some of the material previously covered elsewhere or by omitting certain topics. We have used it on a short course by assuming prior knowledge of most of the material in Chapter 3 and by omitting Chapters 2, 9, and 13. Alternatively, or in addition, Chapter 10 could be omitted. For example, omitting Chapters 2, 9, 10, and 13 leaves sufficient material for a one-semester course of about 22–25 hours, supplemented by five to ten

xiv Preface

tutorials or problem classes. This is not to say that these chapters are unimportant but simply that time constraints sometimes mean not everything can be covered. It is better to work carefully through parts of the book than to attempt to cover it all superficially.

Many people have been involved, either directly or indirectly, in the development of this book. We particularly wish to acknowledge the contribution of Tim Ball, a consultant statistician in Wellington, New Zealand. Tim has freely made material available to us and has been a consistent advocate of the importance of statistical thinking in business. We also wish to acknowledge Jocelyn Dale, Maurice Fletcher, Jock MacKay, Bruce Miller, Peter Mullins, and Peter Scholtes, all of whom have made contributions in various ways. We would also like to thank Kirsty Stroud, formerly our editor at CRC Press, for her encouragement and suggestions on this book. Over the years many undergraduate and MBA students as well as managers and workers in industry have been involved in courses based on all or part of this book. Their feedback and suggestions have been constructive and most welcome. Of course, any mistakes and errors are our responsibility alone.

CONTENTS

re	face .		X1
1	Var	iation	
	1.1	Variation	
	1.2	Airport Immigration	
	1.3	Debt Recovery	
	1.4	Timber Mill Accidents	
	1.5	Budget Deviations	4
	1.6	Red Beads Experiment	4
	1.7	Some Conclusions	12
	1.8	Chapter Summary	13
	1.9	Exercises	13
2	Pro	blem Solving	15
	2.1	A Scientific Approach to Problem Solving	
	2.2	The Need for Data	
	2.3	Deciding What to Measure	17
	2.4	Collecting Data	18
	2.5	Flowcharts and Identifying Causes	19
	2.6	Case Study: Debt Recovery Project	21
	2.7	Chapter Summary	26
	2.8	Exercises	26
3	Loc	oking at Data	29
	3.1	Types of Data	29
	3.2	Presentation of Data	31
	3.3	Rounding Data	32
	3.4	Effective Tables	33
	3.5	Graphs	33

vi

	3.6	Bar Charts	34
	3.7	Pareto Diagrams	36
	3.8	Stem and Leaf Plots	39
	3.9	Box Plots	42
	3.10	Histograms	45
	3.11	Cumulative Frequency Plots	47
		Run Charts	
	3.13	Scatterplots	50
	3.14	Stratification	52
	3.15	Case Study: Aerofoil Effectiveness	55
		Chapter Summary	
	3.17	7 Exercises	58
4	Mod	delling Datadelling Data	67
	4.1	Introduction	67
	4.2	Distributions	70
	4.3	Arithmetic Mean	71
	4.4	Standard Deviation	72
	4.5	Calculating the Mean and Standard Deviation	75
	4.6	Interpreting the Standard Deviation	
	4.7	Summary Statistics	
	4.8	Distribution Shapes	
	4.9	The Normal Distribution	83
	4.10	Case Study: Julia's Diner	
		Binary Data	
		Working with Proportions	
		Working with Counts	
		Case Study: Inter-Island Airways	
		Chapter Summary	
		Exercises	
5	Attr	ribute Data	107
	5.1	Analysing Attribute Data	107
	5.2	Comparing Occurrences between Categories	
	5.3	Excel Demonstration	
	5.4	Analysing a Two-Way Table	
	5.5	Looking at Margins	
	5.6	Case Study: Charnwood Casting Company	121
	5.7	Case Study: Arcas Appliance Company	
	5.8	Chapter Summary	
	5.9	Exercises	
6	Sam	pling	131
	6.1	Collecting Data	131
	6.2	Sampling	

Contents

	6.3	Sampling Concepts	134
	6.4	Wood Experiment	135
	6.5	Random Sampling	141
	6.6	Nonrandom Sampling	144
	6.7	The Accuracy of Sample Estimates	146
	6.8	Chapter Summary	147
	6.9	Data for the Wood Experiment	147
	6.10	Exercises	147
7	Esti	mation	15 3
	7.1	Sampling Error	_
	7.2	Estimating a Mean	
	7.3	Margin of Error for a Mean	
	7.4	Interpreting the Margin of Error	
	7.5	Sample Size for Means	
	7.6	Unknown Population Standard Deviation	
	7.7	Finite Populations	
	7.8	What Affects the Margin of Error?	
	7.9	Estimating a Proportion	
	7.10	Sample Size for Proportions	
		Tests of Hypotheses	
		Case Study: Advertising Awareness	
		Chapter Summary	
		Exercises	
8	Reg	ression Analysis	181
	8.1	Introduction	181
	8.2	Relationships between Variables	
	8.3	Regression Analysis	
	8.4	Least Squares Method	
	8.5	Strength of the Relationship	
	8.6	Analysing the Residuals	
	8.7	Unusual Observations	
	8.8	Prediction from Regression	204
	8.9	Margin of Error	
	8.10	Case Study: Wine Production	
		Chapter Summary	
	8.12	Exercises	211
9	Mul	tiple Regression	219
	9.1	Introduction	
	9.2	Multiple R ²	220
	9.3	Multiple Regression	
	9.4	Polynomial Regression	227
		Using Dummy Variables	

	9.6 Variable Selection and Model Building	233
	9.7 Case Study: Autoparts Ltd.	234
	9.8 Chapter Summary	238
	9.9 Exercises	238
10	Forecasting	249
	10.1 Introduction	249
	10.2 Time Series	251
	10.3 Estimating the Level	254
	10.4 Forecasting with a Trend	259
	10.5 Including Seasonality	263
	10.6 Residual Analysis	268
	10.7 Forecasting Accuracy	268
	10.8 Case Study: Pine Products	270
	10.9 Case Study: Asian Tourism in New Zealand	273
	10.10 Chapter Summary	277
	10.11 Exercises	277
11	Statistical Process Control	287
	11.1 Introduction	287
	11.2 Processes	288
	11.3 Benefits of Reducing Variation	290
	11.4 Common Cause Highway	
	11.5 The 85/15 Rule	
	11.6 Stability and Predictability	
	11.7 Process Adjustment	298
	11.8 Chapter Summary	305
	11.9 Data for the Dice Experiment	305
	11.10 Data for the Quincunx Experiment	305
	11.11 Exercises	306
12	Control Charts	309
	12.1 Introduction	309
	12.2 Three Sigma Limits	310
	12.3 Special Cause Patterns	
	12.4 Variables Control Charts	
	12.5 The Individuals Chart	
	12.6 R and X-Bar Charts	
	12.7 Attribute Charts	
	12.8 Monitoring the Process	
	12.9 Sensitivity of Control Charts	
	12.10 Implementing Statistical Process Control	
	12.11 Case Study: Mike's Orange Company Ltd	
	12.12 Chapter Summary	
	12.13 Exercises	220

13	Imp	ovement Strategies	337
	13.1	Introduction	337
	13.2	Improvement Strategies	
	13.3	Special Cause Strategies	
	13.4	Common Cause Strategies	
	13.5	Specifications	
	13.6	Process Capability	
	13.7	Case Study: The Helicopter Experiment	349
	13.8	Case Study: Cape Corp Debt Collection Agency	354
	13.9	Chapter Summary	
	13.10	Data for the Helicopter Experiment	
	13.11	Exercises	356
Pos	tscrip		367
	_	ions Based on Data	_
		ring Statistically: The Key Principles	
		We Have Not Done	
		tical Computing	
Apr	endix	: Statistical Tables	375
		Normal Distribution	
		ercentage Points of the Normal Distribution	
		ercentage Points of the <i>t</i> -Distribution	
		ercentage Points of the χ^2 -Distribution	
		andom Numbers	
		Constants for Control Charts	
Bib	liogra _l	ohy and Further Reading	383
Ind	ex	***************************************	385

Chapter 1

Variation

1.1 Variation

We live in a variable world. Managers have to make decisions in an environment subject to variability. Without understanding the nature of variability, managers face the danger of making wrong decisions and mistakes that can be very costly to their organisations. Variation exists in all processes, and reducing variation is the key to improving productivity, quality, and profitability. However, as W. Edwards Deming said, "The central problem in management and in leadership... is failure to understand the information in variation." Statistics tells us how to deal with variation, and how to collect and use data so that we can make informed decisions.

In this chapter we discuss four problems involving variability. In each case, we would like you to think carefully about what you would do if you were the manager involved and to compare what you think with what the manager actually did. Then we shall run Deming's red beads experiment and revisit these examples in light of this experiment. The lessons learned should give you insight into the nature of variability and an indication of what should be done.

1.2 Airport Immigration

The data in Figure 1.1 represent the number of passengers processed by immigration officers at an international airport over a typical 1-hour period when the officers would be expected to be quite busy. The passengers processed were those requiring a visa to enter the country. Management expected their officers to process ten passengers during this period. The manager of immigration services, in reviewing these figures, was concerned about the performance of Colin and was thinking about how best to reward Frank. He decided to send a letter to Colin giving him an official warning about his poor performance.

Immigration officer	Number of passengers processed
Alan	9
Barbara	10
Colin	4
Dave	8
Enid	6
Frank	14

Figure 1.1 Customers processed.

He made it clear that unless his results improved, they would have to consider terminating his employment. He then rang Frank and congratulated him on his good work and said that the company was looking at him as future management material.

Q1. Did the manager take appropriate action?

Q2. What would you do, and why?

Q3. Would you recommend anything different if the data in Figure 1.1 gave the number of new cars sold in one week by five salespeople working for Hyden Car Dealership?

1.3 Debt Recovery

In Albion Cleaning, an office cleaning company in New Zealand, the level of unrecovered debt has been a cause of concern for some time. The data in Figure 1.2 give the percentage of invoices that have been paid within the due month. The company is aiming for a target recovery level of 80%. When the amount of recovered debt is much lower than this (percentages in the low to mid-70s, for instance), the general manager visits all the district offices in New Zealand to remind the district managers of the importance of customers paying on time.

Variation 3

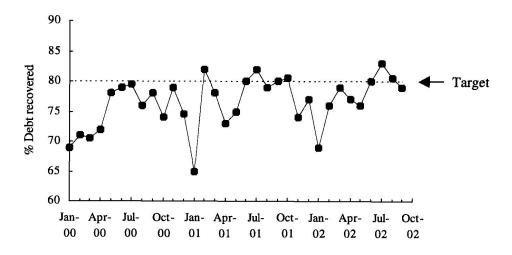


Figure 1.2 Percentage of debts collected within due month.

He tells them their future performance bonuses are at risk unless they quickly reduce the amount of unrecovered debt.

```
Q4. Is the general manager taking appropriate action?
Q5. Has the overall level of debt improved as a result?
Q6. What would you do, and why?
```

1.4 Timber Mill Accidents

Joe Birch, the Managing Director of Greenfields Forestry and Forest Products Ltd., was concerned about the number of accidents occurring in the company's timber mills. In an attempt to reduce the number of accidents, Joe wrote a letter to all timber mill workers informing them that anyone involved in an accident in the future would immediately be dismissed from his or her job.

```
Q7. Has Joe taken appropriate action?

Q8. What would you do, and why?
```

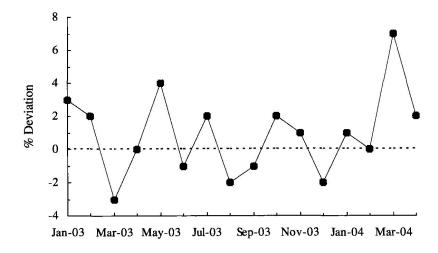


Figure 1.3 Budget deviations.

1.5 Budget Deviations

Jodi Maison, the financial director of Henford Construction Ltd., has been looking at the financial performance of the company over the last year or so. She has collected data on budget deviations, which measure the difference between the amounts budgeted and the actual amounts, expressed as a percentage of the budgeted amounts. The graph in Figure 1.3 gives the budget deviations for a particular account over a 16-month period. The aim, of course, is to have a zero deviation, but some variation is inevitable. Most of the variation lies between –3% and 4%. Jodi was concerned about the 7% deviation in March 2004, so she asked John Allsop, the account manager, to prepare a report explaining why the deviation had been so large that month.

Q9. How should the financial director have reacted to the 7% deviation in March 2004?

1.6 Red Beads Experiment

W. Edwards Deming was a prominent guru of quality management. In 1950 he started to teach the Japanese how to improve quality through the use of statistical quality control methods. His management philosophy had a big impact on industry in the United States and elsewhere from 1980 on. Mary Walton, in her book *The Deming Management Method* (Putnam Publishing, 1986), gives a fascinating account of Deming's methods, including his 14 points of management, and devotes a whole chapter to an important experiment on variability that Deming used to carry out in his teaching: the red beads experiment.

Variation 5

We strongly recommend that you experience the beads experiment in a group situation. We run it in a large class using the equipment described below. A number of people get directly involved in the exercise; others take on a managerial role and tell the workers how they can do better! Apart from being very instructive, it is also a fun exercise. If you do not have the equipment, then an alternative is to use the Excel spreadsheet (*Beads.xls*) provided on the CD-ROM. This will allow you to simulate the results from a typical beads experiment.

In the remainder of this section we first describe how to run the experiment using the equipment. Then we explain how to use the beads spreadsheet. The experiment is then run and the data collected, discussed, and plotted. A discussion of the results follows. Finally, we consider what would happen if the experiment were run a large number of times.

Sometiment Repair Repai

The tools for the experiment are:

- A large number of small plastic beads, the majority of which are white and the others red
- A paddle with 50 bead-size holes in five rows of 10 each, as shown in Figure 1.4
- Two plastic boxes, at least one of which is large enough for dipping the paddle

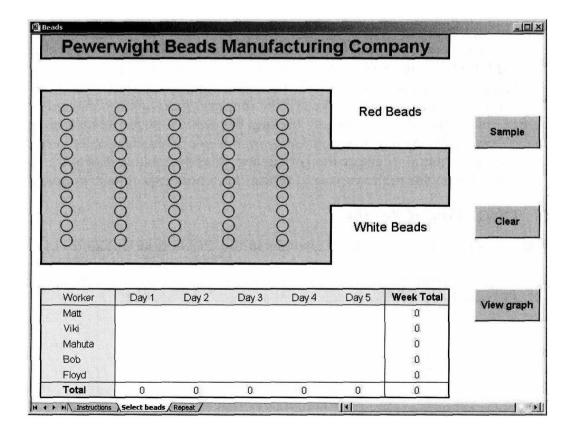


Figure 1.4 Beads experiment spreadsheet.