Statistics and Probability for Engineering Applications

With Microsoft® Excel

Statistics and Probability for Engineering Applications

With Microsoft® Excel

by

W.J. DeCoursey

College of Engineering,

University of Saskatchewan

Saskatoon

Newnes

Amsterdam Boston London New York Oxford Paris
San Diego San Francisco Singapore Sydney Tokyo

Newnes is an imprint of Elsevier Science.

Recognizing the importance of preserving what has been written, Elsevier Science prints its books on acid-free paper whenever possible.

Library of Congress Cataloging-in-Publication Data

ISBN: 0-7506-7618-3

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

The publisher offers special discounts on bulk orders of this book.
For information, please contact:

Manager of Special Sales
Elsevier Science
225 Wildwood Avenue
Woburn, MA 01801-2041
Tel: 781-904-2500
Fax: 781-904-2620

For information on all Newnes publications available, contact our World Wide Web home page at: http://www.newnespress.com

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

Contents

Preface

This book has been written to meet the needs of two different groups of readers. On one hand, it is suitable for practicing engineers in industry who need a better understanding or a practical review of probability and statistics. On the other hand, this book is eminently suitable as a textbook on statistics and probability for engineering students.

Areas of practical knowledge based on the fundamentals of probability and statistics are developed using a logical and understandable approach which appeals to the reader's experience and previous knowledge rather than to rigorous mathematical development. The only prerequisites for this book are a good knowledge of algebra and a first course in calculus. The book includes many solved problems showing applications in all branches of engineering, and the reader should pay close attention to them in each section. The book can be used profitably either for private study or in a class.

Some material in earlier chapters is needed when the reader comes to some of the later sections of this book. Chapter 1 is a brief introduction to probability and statistics and their treatment in this work. Sections 2.1 and 2.2 of Chapter 2 on Basic Probability present topics that provide a foundation for later development, and so do sections 3.1 and 3.2 of Chapter 3 on Descriptive Statistics. Section 4.4, which discusses representing data for a continuous variable in the form of grouped frequency tables and their graphical equivalents, is used frequently in later chapters. Mathematical expectation and the variance of a random variable are introduced in section 5.2. The normal distribution is discussed in Chapter 7 and used extensively in later discussions. The standard error of the mean and the Central Limit Theorem of Chapter 8 are important topics for later chapters. Chapter 9 develops the very useful ideas of statistical inference, and these are applied further in the rest of the book. A short statement of prerequisites is given at the beginning of each chapter, and the reader is advised to make sure that he or she is familiar with the prerequisite material.

This book contains more than enough material for a one-semester or one-quarter course for engineering students, so an instructor can choose which topics to include. Sections on use of the computer can be left for later individual study or class study if so desired, but readers will find these sections using Excel very useful. In my opinion a course on probability and statistics for undergraduate engineering students should include at least the following topics: introduction (Chapter 1), basic probability (sections 2.1 and 2.2), descriptive statistics (sections 3.1 and 3.2), grouped frequency (section 4.4), basics of random variables (sections 5.1 and 5.2), the binomial distribution (section 5.3) (not absolutely essential), the normal distribution (sections 7.1, 7.2, 7.3), variance of sample means and the Central Limit Theorem (from Chapter 8), statistical inferences for the mean (Chapter 9), and regression and correlation (from Chapter 14). A number of other topics are very desirable, but the instructor or reader can choose among them.

It is a pleasure to thank a number of people who have made contributions to this book in one way or another. The book grew out of teaching a section of a general engineering course at the University of Saskatchewan in Saskatoon, and my approach was affected by discussions with the other instructors. Many of the examples and the problems for readers to solve were first suggested by colleagues, including Roy Billinton, Bill Stolte, Richard Burton, Don Norum, Ernie Barber, Madan Gupta, George Sofko, Dennis O'Shaughnessy, Mo Sachdev, Joe Mathews, Victor Pollak, A.B. Bhattacharya, and D.R. Budney. Discussions with Dennis O'Shaughnessy have been helpful in clarifying my ideas concerning the paired t-test and blocking. Example 7.11 is based on measurements done by Richard Evitts. Colleagues were very generous in reading and commenting on drafts of various chapters of the book; these include Bill Stolte, Don Norum, Shehab Sokhansanj, and particularly Richard Burton. Bill Stolte has provided useful comments after using preliminary versions of the book in class. Karen Burlock typed the first version of Chapter 7. I thank all of these for their contributions. Whatever errors remain in the book are, of course, my own responsibility.

I am grateful to my editor, Carol S. Lewis, for all her contributions in preparing this book for publication. Thank you, Carol!

W.J. DeCoursey
Department of Chemical Engineering
College of Engineering
University of Saskatchewan
Saskatoon, SK, Canada
S7N 5A9

What's on the CD-ROM?

Included on the accompanying CD-ROM:

- a fully searchable eBook version of the text in Adobe pdf form
- data sets to accompany the examples in the text
- in the "Extras" folder, useful statistical software tools developed by the Statistical Engineering Division, National Institute of Science and Technology (NIST). Once again, you are cautioned not to apply any technique blindly without first understanding its assumptions, limitations, and area of application.

  Refer to the Read-Me file on the CD-ROM for more detailed information on these files and applications.

List of Symbols

$\overline{A}$ or A'	complement of A
$A \cap B$	intersection of A and B
$A \cup B$	union of A and B
$B \mid A$	conditional probability
$E(X)$	expectation of random variable X
$f(x)$	probability density function
f_i	frequency of result x_i
i	order number
n	number of trials
${}_nC_r$	number of combinations of n items taken r at a time
${}_nP_r$	number of permutations of n items taken r at a time
p	probability of "success" in a single trial
$\hat{p}$	estimated proportion
$p(x_i)$	probability of result x_i
Pr [...]	probability of stated outcome or event
q	probability of "no success" in a single trial
$Q(f)$	quantile larger than a fraction f of a distribution
s	estimate of standard deviation from a sample
s^2	estimate of variance from a sample
s_c^2	combined or pooled estimate of variance
$s_{y\mid x}^2$	estimated variance around a regression line
t	interval of time or space. Also the independent variable of the t-distribution.
X (capital letter)	a random variable
x (lower case)	a particular value of a random variable
$\overline{x}$	arithmetic mean or mean of a sample
z	ratio between $(x - \mu)$ and σ for the normal distribution
α	regression coefficient
β	regression coefficient
λ	mean rate of occurrence per unit time or space
μ	mean of a population
σ	standard deviation of population
$\sigma_{\overline{x}}$	standard error of the mean
σ^2	variance of population

CHAPTER 1

Introduction: Probability and Statistics

Probability and statistics are concerned with events which occur *by chance*. Examples include occurrence of accidents, errors of measurements, production of defective and nondefective items from a production line, and various games of chance, such as drawing a card from a well-mixed deck, flipping a coin, or throwing a symmetrical six-sided die. In each case we may have some knowledge of the likelihood of various possible results, but we cannot predict with any certainty the outcome of any particular trial. Probability and statistics are used throughout engineering. In electrical engineering, signals and noise are analyzed by means of probability theory. Civil, mechanical, and industrial engineers use statistics and probability to test and account for variations in materials and goods. Chemical engineers use probability and statistics to assess experimental data and control and improve chemical processes. It is essential for today's engineer to master these tools.

1.1 Some Important Terms

(a) *Probability* is an area of study which involves predicting the relative likelihood of various outcomes. It is a mathematical area which has developed over the past three or four centuries. One of the early uses was to calculate the odds of various gambling games. Its usefulness for describing errors of scientific and engineering measurements was soon realized. Engineers study probability for its many practical uses, ranging from quality control and quality assurance to communication theory in electrical engineering. Engineering measurements are often analyzed using statistics, as we shall see later in this book, and a good knowledge of probability is needed in order to understand statistics.

(b) *Statistics* is a word with a variety of meanings. To the man in the street it most often means simply a collection of numbers, such as the number of people living in a country or city, a stock exchange index, or the rate of inflation. These all come under the heading of *descriptive statistics*, in which items are counted or measured and the results are combined in various ways to give useful results. That type of statistics certainly has its uses in engineering, and we will deal with it later, but another type of statistics will engage our attention in this book to a much greater extent. That is *inferential statistics* or statistical inference. For example, it is often not practical to measure all the items produced by a process. Instead, we very frequently take a sample and measure the relevant quantity on each member of the sample. We infer something about all the items of interest from our knowledge of the sample. A particular characteristic of all the items we are interested in constitutes a *population*. Measurements of the diameter of all possible bolts as they come off a production process would make up a particular population. A *sample* is a *chosen part* of the population in question, say the measured diameters of twelve bolts chosen to be representative of all the bolts made under certain conditions. We need to know how reliable is the information inferred about the population on the basis of our measurements of the sample. Perhaps we can say that "nineteen times out of twenty" the error will be less than a certain stated limit.

(c) *Chance* is a necessary part of any process to be described by probability or statistics. Sometimes that element of chance is due partly or even perhaps entirely to our lack of knowledge of the details of the process. For example, if we had complete knowledge of the composition of every part of the raw materials used to make bolts, and of the physical processes and conditions in their manufacture, in principle we could predict the diameter of each bolt. But in practice we generally lack that complete knowledge, so the diameter of the next bolt to be produced is an unknown quantity described by a random variation. Under these conditions the distribution of diameters can be described by probability and statistics. If we want to improve the quality of those bolts and to make them more uniform, we will have to look into the causes of the variation and make changes in the raw materials or the production process. But even after that, there will very likely be a random variation in diameter that can be described statistically.

Relations which involve chance are called *probabilistic* or *stochastic* relations. These are contrasted with deterministic relations, in which there is no element of chance. For example, Ohm's Law and Newton's Second Law involve no element of chance, so they are deterministic. However, measurements based on either of these laws do involve elements of chance, so relations between the measured quantities are probabilistic.

(d) Another term which requires some discussion is randomness. A *random* action cannot be predicted and so is due to chance. A *random sample* is one in which every member of the population has an equal likelihood of appearing. Just which items appear in the sample is determined completely by chance. If some items are more likely to appear in the sample than others, then the sample is not random.

1.2 What does this book contain?

We will start with the basics of probability and then cover descriptive statistics. Then various probability distributions will be investigated. The second half of the book will be concerned mostly with statistical inference, including relations between two or more variables, and there will be introductory chapters on design and analysis of experiments. Solved problem examples and problems for the reader to solve will be important throughout the book. The great majority of the problems are directly applied to engineering, involving many different branches of engineering. They show how statistics and probability can be applied by professional engineers.

Some books on probability and statistics use rigorous definitions and many derivations. Experience of teaching probability and statistics to engineering students has led the writer of this book to the opinion that a rigorous approach is not the best plan. Therefore, this book approaches probability and statistics without great mathematical rigor. Each new concept is described clearly but briefly in an introductory section. In a number of cases a new concept can be made more understandable by relating it to previous topics. Then the focus shifts to examples. The reader is presented with carefully chosen examples to deepen his or her understanding, both of the basic ideas and of how they are used. In a few cases mathematical derivations are presented. This is done where, in the opinion of the author, the derivations help the reader to understand the concepts or their limits of usefulness. In some other cases relationships are verified by numerical examples. In still others there are no derivations or verifications, but the reader's confidence is built by comparisons with other relationships or with everyday experience. The aim of this book is to help develop in the reader's mind a clear *understanding* of the ideas of probability and statistics and of the ways in which they are used in practice. The reader must keep the assumptions of each calculation clearly in mind as he or she works through the problems. As in many other areas of engineering, it is *essential* for the reader to do many problems and to understand them thoroughly.

This book includes a number of computer examples and computer exercises which can be done using Microsoft Excel®. Computer exercises are included because statistical calculations from experimental data usually require many repetitive calculations. The digital computer is well suited to this situation. Therefore a book on probability and statistics would be incomplete nowadays if it did not include exercises to be done using a computer. The use of computers for statistical calculations is introduced in sections 3.4 and 4.5.

There is a danger, however, that the reader may obtain only an incomplete understanding of probability and statistics if the fundamentals are neglected in favor of extensive computer exercises. The reader should certainly perform several of the more basic problems in each section before doing the ones which are marked as computer problems. Of course, even the more basic problems can be performed using a spreadsheet rather than a pocket calculator, and that is often desirable. Even if a spreadsheet is used, some of the simpler problems which do not require repetitive