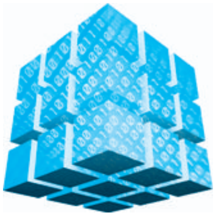

INTRODUCTION TO PROBABILITY AND STATISTICS FOR ENGINEERS AND SCIENTISTS

Fourth Edition

Sheldon M. Ross
Department of Industrial Engineering and Operations Research
University of California, Berkeley

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier

Contents

* Denotes optional material.

Preface

The fourth edition of this book continues to demonstrate how to apply probability theory to gain insight into real, everyday statistical problems and situations. As in the previous editions, carefully developed coverage of probability motivates probabilistic models of real phenomena and the statistical procedures that follow. This approach ultimately results in an intuitive understanding of statistical procedures and strategies most often used by practicing engineers and scientists.

This book has been written for an introductory course in statistics or in probability and statistics for students in engineering, computer science, mathematics, statistics, and the natural sciences. As such it assumes knowledge of elementary calculus.

ORGANIZATION AND COVERAGE

Chapter 1 presents a brief introduction to statistics, presenting its two branches of descriptive and inferential statistics, and a short history of the subject and some of the people whose early work provided a foundation for work done today.

The subject matter of descriptive statistics is then considered in **Chapter 2**. Graphs and tables that describe a data set are presented in this chapter, as are quantities that are used to summarize certain of the key properties of the data set.

To be able to draw conclusions from data, it is necessary to have an understanding of the data's origination. For instance, it is often assumed that the data constitute a "random sample" from some population. To understand exactly what this means and what its consequences are for relating properties of the sample data to properties of the entire population, it is necessary to have some understanding of probability, and that is the subject of **Chapter 3**. This chapter introduces the idea of a probability experiment, explains the concept of the probability of an event, and presents the axioms of probability.

Our study of probability is continued in **Chapter 4**, which deals with the important concepts of random variables and expectation, and in **Chapter 5**, which considers some special types of random variables that often occur in applications. Such random variables as the binomial, Poisson, hypergeometric, normal, uniform, gamma, chi-square, t, and F are presented.

In **Chapter 6**, we study the probability distribution of such sampling statistics as the sample mean and the sample variance. We show how to use a remarkable theoretical result of probability, known as the central limit theorem, to approximate the probability distribution of the sample mean. In addition, we present the joint probability distribution of the sample mean and the sample variance in the important special case in which the underlying data come from a normally distributed population.

Chapter 7 shows how to use data to estimate parameters of interest. For instance, a scientist might be interested in determining the proportion of Midwestern lakes that are afflicted by acid rain. Two types of estimators are studied. The first of these estimates the quantity of interest with a single number (for instance, it might estimate that 47 percent of Midwestern lakes suffer from acid rain), whereas the second provides an estimate in the form of an interval of values (for instance, it might estimate that between 45 and 49 percent of lakes suffer from acid rain). These latter estimators also tell us the "level of confidence" we can have in their validity. Thus, for instance, whereas we can be pretty certain that the exact percentage of afflicted lakes is not 47, it might very well be that we can be, say, 95 percent confident that the actual percentage is between 45 and 49.

Chapter 8 introduces the important topic of statistical hypothesis testing, which is concerned with using data to test the plausibility of a specified hypothesis. For instance, such a test might reject the hypothesis that fewer than 44 percent of Midwestern lakes are afflicted by acid rain. The concept of the *p*-value, which measures the degree of plausibility of the hypothesis after the data have been observed, is introduced. A variety of hypothesis tests concerning the parameters of both one and two normal populations are considered. Hypothesis tests concerning Bernoulli and Poisson parameters are also presented.

Chapter 9 deals with the important topic of regression. Both simple linear regression — including such subtopics as regression to the mean, residual analysis, and weighted least squares — and multiple linear regression are considered.

Chapter 10 introduces the analysis of variance. Both one-way and two-way (with and without the possibility of interaction) problems are considered.

Chapter 11 is concerned with goodness of fit tests, which can be used to test whether a proposed model is consistent with data. In it we present the classical chi-square goodness of fit test and apply it to test for independence in contingency tables. The final section of this chapter introduces the Kolmogorov–Smirnov procedure for testing whether data come from a specified continuous probability distribution.

Chapter 12 deals with nonparametric hypothesis tests, which can be used when one is unable to suppose that the underlying distribution has some specified parametric form (such as normal).

Chapter 13 considers the subject matter of quality control, a key statistical technique in manufacturing and production processes. A variety of control charts, including not only the Shewhart control charts but also more sophisticated ones based on moving averages and cumulative sums, are considered.

Chapter 14 deals with problems related to life testing. In this chapter, the exponential, rather than the normal, distribution plays the key role.

In **Chapter 15** (new to the fourth edition), we consider the statistical inference techniques of bootstrap statistical methods and permutation tests. We first show how probabilities can be obtained by simulation and then how to utilize simulation in these statistical inference approaches.

About the CD

Packaged along with the text is a PC disk that can be used to solve most of the statistical problems in the text. For instance, the disk computes the p-values for most of the hypothesis tests, including those related to the analysis of variance and to regression. It can also be used to obtain probabilities for most of the common distributions. (For those students without access to a personal computer, tables that can be used to solve all of the problems in the text are provided.)

One program on the disk illustrates the central limit theorem. It considers random variables that take on one of the values 0, 1, 2, 3, 4, and allows the user to enter the probabilities for these values along with an integer n. The program then plots the probability mass function of the sum of n independent random variables having this distribution. By increasing n, one can "see" the mass function converge to the shape of a normal density function.

ACKNOWLEDGMENTS

We thank the following people for their helpful comments on the Fourth Edition:

- Charles F. Dunkl, University of Virginia, Charlottesville
- Gabor Szekely, Bowling Green State University
- Krzysztof M. Ostaszewski, Illinois State University
- Michael Ratliff, Northern Arizona University
- Wei-Min Huang, Lehigh University
- Youngho Lee, Howard University
- Jacques Rioux, Drake University
- Lisa Gardner, Bradley University
- Murray Lieb, New Jersey Institute of Technology
- Philip Trotter, Cornell University