时代教育 • 国外高校优秀教材精选

（英文版·原书第6版）

概率与统计（理工类）

Probability and Statistics for Engineering and the Sciences

（美）Jay L. Devore 著

机械工业出版社
CHINA MACHINE PRESS

附赠光盘

时代教育·国外高校优秀教材精选

概率与统计

（理工类）

（英文版·原书第6版）

Probability and Statistics

for Engineering and the Sciences

（美）Jay L. Devore 著

机械工业出版社

Jay L.Devore

Probability and Statistics for Engineering and the Sciences,6e

EISBN: 0-534-39933-9

Copyright © 2004 by Brooks/Cole, a division of Thomson Learning . All rights reserved.

Original language published by Thomson Learning (a division of Thomson Learning Asia Pte Ltd). All Rights reserved. 本书原版由汤姆森学习出版集团出版。版权所有，盗印必究。

China Machine Press is authorized by Thomson Learning to publish and distribute exclusively this English language reprint edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan). Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

本书英文影印版由汤姆森学习出版集团授权机械工业出版社独家出版发行。本版本仅限在中华人民共和国境内（不包括中国香港、澳门特别行政区及中国台湾）销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可，不得以任何方式复制或发行本书的任何部分。

981-254-712-6

北京市版权局著作权合同登记号：图字 01-2004-5162 号

图书在版编目（CIP）数据

概率与统计．理工类：第 6 版/（美）德沃尔（Devore, J. L. ）著.－北京：机械工业出版社，2005.1

（时代教育·国外高校优秀教材精选）

ISBN 7-111-15724-9

Ⅰ．概… Ⅱ．德… Ⅲ．①概率论－高等学校－教材－英文 ②数理统计－高等学校－教材－英文 Ⅳ．021

中国版本图书馆 CIP 数据核字（2004）第 125137 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）

责任编辑：张祖凤

封面设计：饶 薇 责任印制：石 冉

三河市宏达印刷有限公司印刷·新华书店北京发行所发行

2005 年 2 月第 1 版第 1 次印刷

787mm×1 092mm 1/16·46.5 印张·1150 千字

定价：72.00 元 （含 1CD）

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

本社购书热线电话（010）68993821、88379646

68326294、68320718

封面无防伪标均为盗版

国外高校优秀教材审定委员会

主任委员：

杨叔子

委员（按姓氏笔画为序）：

丁丽娟	王先逵	王大康	白峰杉	石德珂
史荣昌	孙洪祥	朱孝禄	陆启韶	张润琦
张　策	张三慧	张福润	张延华	吴宗泽
吴　麒	宋心琦	李俊峰	佘远斌	陈文楷
陈立周	单辉祖	俞正光	赵汝嘉	郭可谦
翁海珊	龚光鲁	章栋恩	黄永畅	谭泽光
郭鸿志				

出 版 说 明

随着我国加入 WTO，国际间的竞争越来越激烈，而国际间的竞争实际上也就是人才的竞争、教育的竞争。为了加快培养具有国际竞争力的高水平技术人才，加快我国教育改革的步伐，国家教育部近来出台了一系列倡导高校开展双语教学、引进原版教材的政策。以此为契机，机械工业出版社陆续推出了一系列国外影印版教材，其内容涉及高等学校公共基础课，以及机、电、信息领域的专业基础课和专业课。

引进国外优秀原版教材，在有条件的学校推动开展英语授课或双语教学，自然也引进了先进的教学思想和教学方法，这对提高我国自编教材的水平，加强学生的英语实际应用能力，使我国的高等教育尽快与国际接轨，必将起到积极的推动作用。

为了做好教材的引进工作，机械工业出版社特别成立了由著名专家组成的国外高校优秀教材审定委员会。这些专家对实施双语教学做了深入细致的调查研究，对引进原版教材提出了许多建设性意见，并慎重地对每一本将要引进的原版教材一审再审，精选再精选，确认教材本身的质量水平，以及权威性和先进性，以期所引进的原版教材能适应我国学生的外语水平和学习特点。在引进工作中，审定委员会还结合我国高校教学课程体系的设置和要求，对原版教材的教学思想和方法的先进性、科学性严格把关。同时尽量考虑原版教材的系统性和经济性。

这套教材出版后，我们将根据各高校的双语教学计划，举办原版教材的教师培训，及时地将其推荐给各高校选用。希望高校师生在使用教材后及时反馈意见和建议，使我们更好地为教学改革服务。

机械工业出版社

此为试读，需要完整PDF请访问：www.ertongbook.com

序

本书原版已是第 6 版，出版于 2004 年，是为理工科本科生、研究生学习而编写的有关统计推理的理论、思想、方法的教材。内容包括了理工科概率论的入门，又包括了经典统计和现代统计的基础部分。具体包括：描述性统计量；概率；离散随机变量与概率分布；连续型随机变量与概率分布；联合分布与随机样本；点估计；单样本区间估计；单样本假设检验；二样本推断；方差分析；多因子方差分析；简单线性回归和相关；非线性与多变量回归分析；类别数据分析与拟合度检验等。

本书具有以下特点：

1）目的性明确，其概率部分是完全为统计而设置的，因此比国内的一般教材要少。要求的数学基础较低，只要具备初等微积分的知识即可通览全书。

2）通过正文与习题旁征博引，引进了一些近代统计处理的新的技术。

3）配有大量的例子与丰富的练习题。

4）书中统计部分与国内的一般教材相比，理论上要深一些，论述的模型更多，案例的涉及面更广，实用面更丰富，统计思想的阐述与算法更具体些。

根据国内的教学实际，本次影印时删去了原书中的第 15、16 章“不依赖分布的方法”及“质量控制方法”的内容。

本书适合作为理工科各专业有关概率统计的双语教学的教材或教学参考书，也可供经济管理类专业教学参考。

清华大学

茆诗松

Preface

Purpose

The use of probability models and statistical methods for analyzing data has become common practice in virtually all scientific disciplines. This book attempts to provide a comprehensive introduction to those models and methods most likely to be encountered and used by students in their careers in engineering and the natural sciences. Although the examples and exercises have been designed with scientists and engineers in mind, most of the methods covered are basic to statistical analyses in many other disciplines, so that students of business and the social sciences will also profit from reading the book.

Approach

Students in a statistics course designed to serve other majors may be initially skeptical of the value and relevance of the subject matter, but my experience is that students *can* be turned on to statistics by the use of good examples and exercises that blend their everyday experiences with their scientific interests. Consequently, I have worked hard to find examples of real, rather than artificial, data—data that someone thought was worth collecting and analyzing. Many of the methods presented, especially in the later chapters on statistical inference, are illustrated by analyzing data taken from a published source, and many of the exercises also involve working with such data. Sometimes the reader may be unfamiliar with the context of a particular problem (as indeed I often was), but I have found that students are more attracted by real problems with a somewhat strange context than by patently artificial problems in a familiar setting.

Mathematical Level

The exposition is relatively modest in terms of mathematical development. Substantial use of the calculus is made only in Chapter 4 and parts of Chapters 5 and 6. In particular, with the exception of an occasional remark or aside, calculus appears in the inference part of the book only in the second section of Chapter 6. Matrix algebra is not used at all. Thus almost all the exposition should be accessible to those whose mathematical background includes one semester or two quarters of differential and integral calculus.

Content

Chapter 1 begins with some basic concepts and terminology—population, sample, descriptive and inferential statistics, enumerative versus analytic studies, and so on—and continues with a survey of important graphical and numerical descriptive methods. A rather traditional development of probability is given in Chapter 2, followed by probability distributions of discrete and continuous random variables in Chapters 3 and 4, respectively. Joint distributions and their properties are discussed in the first part of Chapter 5. The latter part of this chapter introduces statistics and their sampling distributions, which form the bridge between probability and inference. The next three chapters cover point estimation, statistical intervals, and hypothesis testing based on a single sample. Methods of inference involving two independent samples and paired data are presented in Chapter 9. The analysis of variance is the subject of Chapters 10 and 11 (single-factor and multifactor, respectively). Regression makes its initial appearance in Chapter 12 (the simple linear regression model and correlation) and returns for an extensive encore in Chapter 13. The last three chapters develop chi-squared methods, distribution-free (nonparametric) procedures, and techniques from statistical quality control.

Helping Students Learn

Although the book's mathematical level should give most science and engineering students little difficulty, working toward an understanding of the concepts and gaining an appreciation for the logical development of the methodology may sometimes require substantial effort. To help students gain such an understanding and appreciation, I have provided numerous exercises ranging in difficulty from many that involve routine application of text material to some that ask the reader to extend concepts discussed in the text to somewhat new situations. There are many more exercises than most instructors would want to assign during any particular course, but I recommend that students be required to work a substantial number of them; in a problem-solving discipline, active involvement of this sort is the surest way to identify and close the gaps in understanding that inevitably arise. Answers to odd-numbered exercises appear in the answer section at the back of the text. In addition, a Student Solutions Manual, consisting of worked-out solutions to virtually all the odd-numbered exercises, is available.

New for This Edition

Changes for this Sixth Edition include the following:

- Many new examples and exercises are included, based of course mostly on real data. Some of these involve less structure than typical textbook exercises. For example, an exercise in Chapter 13 gives a multiple regression data set (which at first blush looks like a simple linear regression data set, except that information about a categorical predictor is also included); the student is simply asked to fit several models and draw appropriate conclusions. An exercise at the end of Chapter 14 includes parts relating to correlation, ANOVA, a two-sample t analysis, and a chi-squared analysis—without specifically saying so. Emphasis is, generally, not only on the mechanics but also

on interpretation. For example, in Chapter 1 the student is asked not only to construct a boxplot, but also to say what it suggests. Some exercises introduce new concepts or techniques, for example, a half-normal plot in Chapter 4 and autocorrelation for a time series in Chapter 12.

- An enlarged definition of a hypothesis at the beginning of Chapter 8 includes an assertion that the population distribution has a particular form (rather than hypotheses being only claims about parameters).
- A more detailed description of pooled t procedures to provide better balance between unpooled and pooled analyses.
- More information is provided about alternatives to traditional formulas for obtaining interval estimates of a single proportion and a difference between proportions.
- The section on multiple comparisons in ANOVA has been completely rewritten to better emphasize the nature of simultaneous confidence level and Tukey's procedure.
- A discussion of MINITAB®'s power calculations in single-factor ANOVA has been added.
- There is more emphasis on the nature of variation in the slope estimate in simple linear regression.
- The LOWESS method for fitting a nonlinear regression function is introduced.
- An alternative interpretation of regression coefficients in multiple regression via partial residual plots is provided.
- Overall, there has been a tightening and fine-tuning of the exposition wherever appropriate.
- A Student's Suite CD includes Java™ applets created by Gary McClelland, specifically for this calculus based text; in addition, the CD will include data sets from the text.
- BCA testing, a test creation and course management tool.

Recommended Coverage

There is enough material in this book for a year-long course. Anyone teaching a course of shorter duration will have to be selective in the choice of topics to be included. At Cal Poly, we teach a two-quarter sequence, meeting four hours per week. During the first ten weeks we cover much of the material in Chapters 1–7 (going lightly over joint distributions and the details of estimation by maximum likelihood and the method of moments). The second quarter begins with hypothesis testing and moves on to two-sample inferences, ANOVA, regression, and selections from the chi-squared, distribution-free, and quality control chapters. Coverage of material in a one-semester course would obviously have to be somewhat more restrictive. There is, of course, never enough time to teach students all that we would like them to know!

Acknowledgments

My colleagues here at Cal Poly have provided me with invaluable support and encouragement over the years. I am also grateful to the many users of previous editions who have made suggestions for improvement (and pointed out occasional errors). A note of thanks goes to Julie Seely for her work on the Student Solutions Manual.

I gratefully acknowledge the plentiful feedback provided by the following reviewers of this and previous editions: Robert L. Armacost, University of Central Florida;

Douglas M. Bates, University of Wisconsin–Madison; David Clark, California State Polytechnic University at Pomona; David M. Cresap, University of Portland; Don E. Deal, University of Houston; Charles E. Donaghey, University of Houston; Patrick J. Driscoll, U.S. Military Academy; Mark Duva, University of Virginia; Nasser S. Fard, Northeastern University; Celso Grebogi, University of Maryland; James J. Halavin, Rochester Institute of Technology; Jennifer Hoeting, Colorado State University; Wei-Min Huang, Lehigh University; Stephen Kokoska, Colgate University; Arnold R. Miller, University of Denver; John J. Millson, University of Maryland; Don Ridgeway, North Carolina State University; Larry J. Ringer, Texas A&M University; Richard M. Soland, The George Washington University; Clifford Spiegelman, Texas A&M University; Jery Stedinger, Cornell University; David Steinberg, Tel Aviv University; G. Geoffrey Vining, University of Florida; Bhutan Wadhwa, Cleveland State University; and Michael G. Zabetakis, University of Pittsburgh.

I very much appreciate the production services provided by Martha Emry (who was always gracious, supportive, and timely in her conversations with me). Thanks also to copy editor Pamela Rockwell. The staff at Duxbury and Brooks/Cole has as usual been extremely supportive—thanks in particular to Carolyn Crockett, Curt Hinrichs, Ann Day, Julie Bliss, Joseph Rogove, and Karen Haga. I wish also to commend the sales representatives of Thomson Learning for their hard work over the years in getting the word out about earlier editions of this book as well as other books I have written. Finally, words cannot adequately express my gratitude toward my wife, Carol, for her support of my writing efforts over the course of the last three decades.

Jay L. Devore

Contents

出版说明
序

1 Overview and Descriptive Statistics 1

Introduction 1
1.1 Populations, Samples, and Processes 3
1.2 Pictorial and Tabular Methods in Descriptive Statistics 11
1.3 Measures of Location 28
1.4 Measures of Variability 36
Supplementary Exercises 48
Bibliography 51

2 Probability 52

Introduction 52
2.1 Sample Spaces and Events 53
2.2 Axioms, Interpretations, and Properties of Probability 58
2.3 Counting Techniques 67
2.4 Conditional Probability 75
2.5 Independence 86
Supplementary Exercises 92
Bibliography 96

3 Discrete Random Variables and Probability Distributions 97

Introduction 97
3.1 Random Variables 98
3.2 Probability Distributions for Discrete Random Variables 101
3.3 Expected Values of Discrete Random Variables 111
3.4 The Binomial Probability Distribution 120
3.5 Hypergeometric and Negative Binomial Distributions 128
3.6 The Poisson Probability Distribution 135
Supplementary Exercises 140
Bibliography 144

此为试读,需要完整PDF请访问: www.ertongbook.com

4 Continuous Random Variables and Probability Distributions 145

Introduction 145
4.1 Continuous Random Variables and Probability Density Functions 146
4.2 Cumulative Distribution Functions and Expected Values 152
4.3 The Normal Distribution 160
4.4 The Gamma Distribution and Its Relatives 174
4.5 Other Continuous Distributions 181
4.6 Probability Plots 188
Supplementary Exercises 199
Bibliography 204

5 Joint Probability Distributions and Random Samples 205

Introduction 205
5.1 Jointly Distributed Random Variables 206
5.2 Expected Values, Covariance, and Correlation 219
5.3 Statistics and Their Distributions 225
5.4 The Distribution of the Sample Mean 237
5.5 The Distribution of a Linear Combination 243
Supplementary Exercises 249
Bibliography 252

6 Point Estimation 253

Introduction 253
6.1 Some General Concepts of Point Estimation 254
6.2 Methods of Point Estimation 269
Supplementary Exercises 278
Bibliography 279

7 Statistical Intervals Based on a Single Sample 281

Introduction 281
7.1 Basic Properties of Confidence Intervals 282
7.2 Large-Sample Confidence Intervals for a Population Mean and Proportion 291
7.3 Intervals Based on a Normal Population Distribution 299
7.4 Confidence Intervals for the Variance and Standard Deviation of a Normal Population 308
Supplementary Exercises 311
Bibliography 314

8 Tests of Hypotheses Based on a Single Sample 315

Introduction 315
8.1 Hypotheses and Test Procedures 316

8.2 Tests About a Population Mean 326
8.3 Tests Concerning a Population Proportion 339
8.4 *P*-Values 344
8.5 Some Comments on Selecting a Test Procedure 352
Supplementary Exercises 356
Bibliography 359

9 Inferences Based on Two Samples 354

Introduction 354
9.1 *z* Tests and Confidence Intervals for a Difference Between Two Population Means 361
9.2 The Two-Sample *t* Test and Confidence Interval 372
9.3 Analysis of Paired Data 381
9.4 Inferences Concerning a Difference Between Population Proportions 391
9.5 Inferences Concerning Two Population Variances 399
Supplementary Exercises 403
Bibliography 409

10 The Analysis of Variance 410

Introduction 410
10.1 Single-Factor ANOVA 411
10.2 Multiple Comparisons in ANOVA 422
10.3 More on Single-Factor ANOVA 428
Supplementary Exercises 438
Bibliography 440

11 Multifactor Analysis of Variance 441

Introduction 441
11.1 Two-Factor ANOVA with $K_{ij} = 1$ 442
11.2 Two-Factor ANOVA with $K_{ij} > 1$ 456
11.3 Three-Factor ANOVA 465
11.4 2^p Factorial Experiments 476
Supplementary Exercises 491
Bibliography 494

12 Simple Linear Regression and Correlation 496

Introduction 496
12.1 The Simple Linear Regression Model 497
12.2 Estimating Model Parameters 505
12.3 Inferences About the Slope Parameter β_1 520
12.4 Inferences Concerning $\mu_{Y \cdot x^*}$ and the Prediction of Future *Y* Values 530
12.5 Correlation 538
Supplementary Exercises 549
Bibliography 554

13 Nonlinear and Multiple Regression 555

Introduction 555
13.1 Aptness of the Model and Model Checking 556
13.2 Regression with Transformed Variables 564
13.3 Polynomial Regression 576
13.4 Multiple Regression Analysis 587
13.5 Other Issues in Multiple Regression 612
Supplementary Exercises 626
Bibliography 632

14 Goodness-of-Fit Tests and Categorical Data Analysis 633

Introduction 633
14.1 Goodness-of-Fit Tests When Category Probabilities Are Completely Specified 634
14.2 Goodness-of-Fit Tests for Composite Hypotheses 642
14.3 Two-Way Contingency Tables 655
Supplementary Exercises 663
Bibliography 666

Appendix Tables 667

A.1 Cumulative Binomial Probabilities 668
A.2 Cumulative Poisson Probabilities 671
A.3 Standard Normal Curve Areas 672
A.4 The Incomplete Gamma Function 674
A.5 Critical Values for t Distributions 675
A.6 Tolerance Critical Values for Normal Population Distributions 676
A.7 Critical Values for Chi-Squared Distributions 677
A.8 t Curve Tail Areas 678
A.9 Critical Values for F Distributions 680
A.10 Critical Values for Studentized Range Distributions 686
A.11 Chi-Squared Curve Tail Areas 687
A.12 Critical Values for the Ryan–Joiner Test of Normality 689
A.13 Critical Values for the Wilcoxon Signed-Rank Test 690
A.14 Critical Values for the Wilcoxon Rank-Sum Test 691
A.15 Critical Values for the Wilcoxon Signed-Rank Interval 692
A.16 Critical Values for the Wilcoxon Rank-Sum Interval 693
A.17 β Curves for t Tests 694

Answers to Odd-Numbered Exercises 695
Index 715

1

Overview and Descriptive Statistics

Introduction

Statistical concepts and methods are not only useful but indeed often indispensable in understanding the world around us. They provide ways of gaining new insights into the behavior of many phenomena that you will encounter in your chosen field of specialization in engineering or science.

The discipline of statistics teaches us how to make intelligent judgments and informed decisions in the presence of uncertainty and variation. Without uncertainty or variation, there would be little need for statistical methods or statisticians. If every component of a particular type had exactly the same lifetime, if all resistors produced by a certain manufacturer had the same resistance value, if pH determinations for soil specimens from a particular locale gave identical results, and so on, then a single observation would reveal all desired information.

An interesting manifestation of variation arises in the course of performing emissions testing on motor vehicles. The expense and time requirements of the Federal Test Procedure (FTP) preclude its widespread use in vehicle inspection programs. As a result, many agencies have developed less costly and quicker tests, which it is hoped replicate FTP results. According to the journal article "Motor Vehicle Emissions Variability" (*J. of the Air and Waste Mgmt. Assoc.*, 1996: 667–675), the acceptance of the FTP as a gold

standard has led to the widespread belief that repeated measurements on the same vehicle would yield identical (or nearly identical) results. The authors of the article applied the FTP to seven vehicles characterized as "high emitters." Here are the results for one such vehicle:

HC (gm/mile)	13.8	18.3	32.2	32.5
CO (gm/mile)	118	149	232	236

The substantial variation in both the HC and CO measurements casts considerable doubt on conventional wisdom and makes it much more difficult to make precise assessments about emissions levels.

How can statistical techniques be used to gather information and draw conclusions? Suppose, for example, that a materials engineer has developed a coating for retarding corrosion in metal pipe under specified circumstances. If this coating is applied to different segments of pipe, variation in environmental conditions and in the segments themselves will result in more substantial corrosion on some segments than on others. Methods of statistical analysis could be used on data from such an experiment to decide whether the *average* amount of corrosion exceeds an upper specification limit of some sort or to predict how much corrosion will occur on a single piece of pipe.

Alternatively, suppose the engineer has developed the coating in the belief that it will be superior to the currently used coating. A comparative experiment could be carried out to investigate this issue by applying the current coating to some segments of pipe and the new coating to other segments. This must be done with care lest the wrong conclusion emerge. For example, perhaps the average amount of corrosion is identical for the two coatings. However, the new coating may be applied to segments that have superior ability to resist corrosion and under less stressful environmental conditions compared to the segments and conditions for the current coating. The investigator would then likely observe a difference between the two coatings attributable not to the coatings themselves, but just to extraneous variation. Statistics offers not only methods for analyzing the results of experiments once they have been carried out but also suggestions for how experiments can be performed in an efficient manner to mitigate the effects of variation and have a better chance of producing correct conclusions.

1.1 Populations, Samples, and Processes

Engineers and scientists are constantly exposed to collections of facts, or **data**, both in their professional capacities and in everyday activities. The discipline of statistics provides methods for organizing and summarizing data and for drawing conclusions based on information contained in the data.

An investigation will typically focus on a well-defined collection of objects constituting a **population** of interest. In one study, the population might consist of all gelatin capsules of a particular type produced during a specified period. Another investigation might involve the population consisting of all individuals who received a B.S. in engineering during the most recent academic year. When desired information is available for all objects in the population, we have what is called a **census**. Constraints on time, money, and other scarce resources usually make a census impractical or infeasible. Instead, a subset of the population—a **sample**—is selected in some prescribed manner. Thus we might obtain a sample of bearings from a particular production run as a basis for investigating whether bearings are conforming to manufacturing specifications, or we might select a sample of last year's engineering graduates to obtain feedback about the quality of the engineering curricula.

We are usually interested only in certain characteristics of the objects in a population: the number of flaws on the surface of each casing, the thickness of each capsule wall, the gender of an engineering graduate, the age at which the individual graduated, and so on. A characteristic may be categorical, such as gender or type of malfunction, or it may be numerical in nature. In the former case, the *value* of the characteristic is a category (e.g., female or insufficient solder), whereas in the latter case, the value is a number (e.g., age = 23 years or diameter = .502 cm). A **variable** is any characteristic whose value may change from one object to another in the population. We shall initially denote variables by lowercase letters from the end of our alphabet. Examples include

x = brand of calculator owned by a student

y = number of major defects on a newly manufactured automobile

z = braking distance of an automobile under specified conditions

Data results from making observations either on a single variable or simultaneously on two or more variables. A **univariate** data set consists of observations on a single variable. For example, we might determine the type of transmission, automatic (A) or manual (M), on each of ten automobiles recently purchased at a certain dealership, resulting in the categorical data set

M A A A M A A M A A

The following sample of lifetimes (hours) of brand D batteries put to a certain use is a numerical univariate data set:

5.6 5.1 6.2 6.0 5.8 6.5 5.8 5.5

We have **bivariate** data when observations are made on each of two variables. Our data set might consist of a (height, weight) pair for each basketball player on a team, with the first observation as (72, 168), the second as (75, 212), and so on. If an engineer

此为试读,需要完整PDF请访问: www.ertongbook.com