

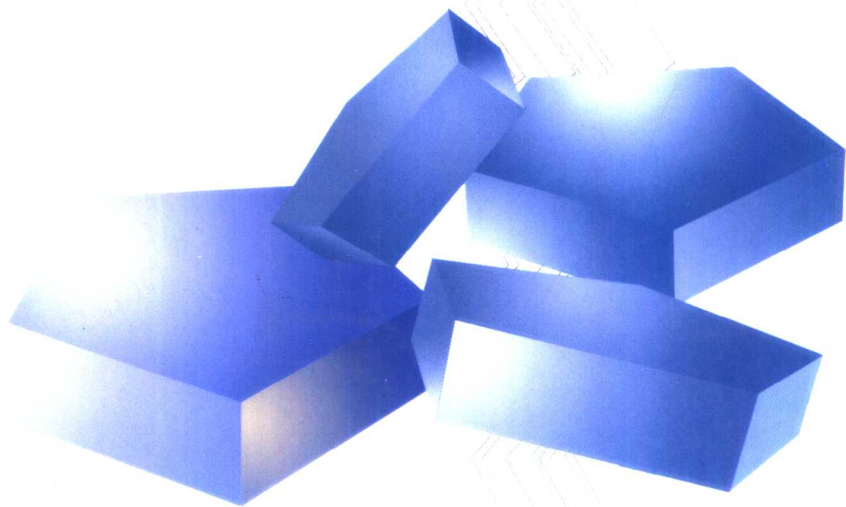
Fundamentals of Biostatistics

# 生物统计学基础

(原书第五版)

[美] 伯纳德·罗斯纳 著

● 孙尚拱 译



 科学出版社  
www.sciencep.com

图字: 01-2002-1679 号

## 内 容 简 介

本书是国外优秀教材 *Fundamentals of Biostatistics* (第五版) 的中译本, 由哈佛大学具有丰富教学经验的一流教授编写。

本书是介绍生物统计学重要知识和基本应用的导论性教材。书中运用丰富的医学和生物学实例及流程图, 生动形象地阐明了生物统计学的概念内涵和方法公式。为了便于读者自学, 本书尽量贯穿初等数学讨论, 而不过多涉及高等数学证明, 并且每章末附摘要、练习题和参考文献, 书末有习题解答、索引及数据光盘。

本书适用于高等院校生物学和医学相关专业师生。

Fundamentals of Biostatistics, 5th edition

By Bernard Rosner

Copyright © 2000 by Brooks/Cole, a division of Thomson Learning

First published by Duxbury Press, a division of Thomson Learning

All Rights Reserved

Authorized Translation/Adaptation of the edition by Thomson Learning and Science Press. No part of this book may be reproduced in any form without the express written permission of Thomson Learning and Science Press.

### 图书在版编目 (CIP) 数据

生物统计学基础/ (美) 罗斯纳 (Rosner, B.) 著; 孙尚拱译. —北京: 科学出版社, 2004

书名原文: Fundamentals of Biostatistics, 5th Edition

ISBN 7-03-010394-7

I. 生… II. ①罗…②孙… III. 生物统计-理论-教材 IV. Q332

中国版本图书馆 CIP 数据核字 (2002) 第 027340 号

责任编辑: 周 辉 杨 波 盖 宇 黄 方/责任校对: 柏连海

责任印制: 安春生/封面设计: 耕者工作室

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

双 青 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

\*

2004年4月第 一 版 开本: B5(720×1000)

2004年4月第一次印刷 印张: 50 1/4

印数: 1—3 000 字数: 985 000

定 价: 65.00 元 (配 盘)

(如有印装质量问题, 我社负责调换〈环伟〉)

## 译 序

译者从事医学统计教学及应用研究 30 余年，读过的国内外统计教材不在少数，而这本医学统计教材，我认为是我所见过的最好的，称之为典范之作也不为过。

这是一本非常有特色的书，与其他同类书比较，其优点非常多。今只讲我认为最重要的三点：

一、这是一本通俗易懂但又非常严谨、细致、深入而又全面的教材。书中的多数公式都有证明，且多用初等数学方法，对于不能用初等方法证明的也都给出出处。对比国内的医学统计教材，其中公式基本上是既没有证明又无出处。

二、本书核心是医学应用，作者通过大量的医学实例，引入及介绍统计方法，从如何构思（第 1 章）到分析结果的解释，几乎都有全过程。由于一切从实用出发，所以对实用极为重要的知识，比如功效（power）及样本量的估计，都是实际工作者极为关心的，至今译者未看到有一本同类书能像本书那样详细地介绍功效及样本量的估计。本书有 351 个医学实例，1166 个练习题，内容几乎涉及医学的所有领域。仔细读这些例子的提出及解决，对提高科研能力极有帮助。

三、学术上的先进性。本书是第 5 版，作者能在每个版本中把当时国际上最流行的统计方法及统计软件及时吸收在书内，本书的第 13 章（流行病学研究中的设计与分析技术）及第 14 章（假设检验：人-时间数据）最为精彩，也最有代表性。

译者译本书时，常有联想，中国的大学教材几乎都不比美国的差，但为什么中国的医学统计教材与美国哈佛大学医学教材差距如此之大？当然与学制不同很有关系，但医学统计专业的研究生又如何？我以为，我们的医学统计教师及流行病学教师们，如都认真地阅读此书，我国的医学统计教学及科研水平必定会有大的提高。

最后，译者非常感谢中国科学院院士陈希孺教授推荐我承担本书翻译工作，也感谢我的同事何平平老师帮我翻译第 6 章及其后的练习题和所有的流程图。

孙尚拱

# 前 言

这本导论性质的生物统计学教材，是为高年级大学生或对医学卫生领域有兴趣的研究生而写的。读者不需要有统计学的背景，他（她）仅需有代数知识即可。

《生物统计学基础》一书是由我在哈佛大学（Harvard University）及哈佛医学院执教 25 年多的教学笔记发展而成。我写本书的目的，是为了帮助学生掌握大量出现在医学文献中的统计方法。从学生角度看，书中引用的大量实际存在于文献中的例子是很有代表性的，它对于理解及掌握统计方法是重要的。因此，这些例子及练习题是本书的基础，它们要么取自医学文献中的实际论文，要么是作者在哈佛医学院从事医学研究及咨询工作中遇到的实际问题。

## 方法

大多数入门的统计教材要么完全不用数学公式，要么采用严密复杂的数学公式。本书介于两者之间，用最少量的数学公式并对重要概念给以完全的解释。本书中的每一个新概念都是从当前医学研究问题的例子中有步骤地引出的。另外，我也引用电脑输出的结果，这对说明这些概念是合适的。

首先，我把本书写成介绍性的生物统计教材。但是这个领域在过去的几年中发展很快，因为新的统计软件包的增加，使我们可以比过去更好地完成复杂的数据分析。由此，本书的第二个目标是在介绍性的水平上去展示新的统计方法，使学生能通晓它们而无需涉足特殊的（通常是更高级）统计教材。

为了更清楚地区分这两个目标，我把介绍性的内容放在前 12 章，它构成了本书的主要部分。然后增加“流行病研究中的设计与分析技术”作为第 13 章，“假设检验：人-时间数据”作为第 14 章。这两章包括了最新流行病学研究中的更高级的统计技术。

## 第 5 版中的变动

在第 5 版中，增加了 11 节新内容并基本上重写了其他 8 节。新的内容如下：

■ 扩充了真实数据的计算机练习题。这个版本中包含有 23 组数据，它们都存放在本书背面的一个软盘中。读者首先应检查这些数据集，它们是用 Excel 可读格式存放的，而在过去的版本中，是用 MINITAB<sup>®</sup> 及 ASCII 格式存放的。

- 一个附加的病例研究——有关抽烟对于骨密度的影响——贯穿于本书的几章。
- 新的或扩充了下面的主题：
  - ROC 曲线 (3.7.1 节)
  - 在微软 Excel 软件上使用电子表格 (4.8.2 节)
  - 协方差 (5.6.1 节及 11.7 节)
  - 随机变量是应变变量时的线性组合的方差及期望值 (5.6.1 节)
  - 使用 Excel 软件完成假设检验及求取置信区间 (第 6 至第 14 章)
  - 纵向研究中样本量的估计 (8.11 节)
  - delta 方法 (13.3 节)
  - 等效性研究 (13.9 节)
  - 再分析 (meta-analysis, 13.8 节)
  - 方差-协方差矩阵 (12.5 节)
  - 估计相关系数时的样本量 (11.8.4 节)
  - 聚集性的二态数据 (13.11 节)
  - 测度误差的方法 (13.12 节)
  - 对于“人-时间”数据中功效及样本量的估计 (14.4 及 14.6 节)
  - 生存分析中样本量的估计 (14.12 节)
  - 对于单样本发病率数据的统计推断 (14.2 节)
  - 发病率的置信区间 (14.2 节)

新的或扩充了的部分，在本版目录中用“\*”加以注明。

## 练习题

本版中包含有 1166 个练习题（而以前的版本中有 893 个），练习题的所有数据也包括在内。每章末尾的习题中带有“\*”的在本书末都给出简单的答案。按照学生的要求，约有 600 个较为复杂的练习题的解答是在 *Study Guide to Accompany Fundamentals of Biostatistics*, 5th edition (ISBN 0-534-37120-5) 中\*。该学习指导提供每一问题的完全解答。另外，上述练习题中约 100 个是很有特色的混合题，它们是随机编排的，并不与书中所在的章有联系。这可以给学生一次锻炼的机会，以决定在什么情形下应该用什么方法。最后，在学习指导的附录中，对于本书中的 Excel 统计命令作了一个简单的描述。

\* 此题解书未包括在本次翻译计划内。——译者注

## 计算方法

计算方法类似于第四版。所有中间结果都是完全精确的（10位以上有意义的数字），即使在书中它们仅以较少的位数（通常是2或3位）出现也是如此。在某些情形中，中间结果似乎与最后结果不一致，但事实并非如此。

## 编制法

第5版的生物统计学基础是按下面形式编写的。

第1章是绪言，把我参与的实际医学研究工作的发展作了一个简述。它展示的生物统计学在医学研究中的作用给人感觉独特。

第2章介绍描述性统计。为了表示医学数据，作者介绍了所有重要的数值及图形工具。这一章对于医学论文的作者及读者都是特别地重要，因为这两者之间的信息传递是通过这些描述性工作而完成的。

第3至第5章介绍概率。内有概率的基本原理，某些最重要的概率分布，比如二项分布及正态分布，这些分布大量地出现于本书后面的章节之中。

第6至第10章介绍统计推理的某些基本方法。

第6章介绍从总体中抽取随机样本的概念。叙述了抽样分布中难弄的记号，并介绍了常见的抽样分布\*，比如 $t$ 分布及卡方分布。另外还有估计理论的基本方法，包括置信区间的拓广性讨论。

第7及第8章包含假设检验的基本原理。内有正态分布数据的最基本的假设检验，比如 $t$ 检验，对单样本及两样本问题作了充分地讨论。

第9章介绍非参数统计的基本原理。放松了对正态分布的假设，从而发展与第7及第8章相似的检验公式而又没有规定分布类型。

第10章介绍用于类型数据的假设检验的基本概念，包括一些最广泛使用的统计过程，比如卡方检验及Fisher精确检验。

第11章发展了回归分析原理。内有简单线性回归及多重回归。还包括回归模型的拟合优度的重要部分，最后介绍秩相关。

第12章介绍方差分析（ANOVA）原理。讨论固定方差的单向分析及随机效应模型。另外，也讨论双向ANOVA及协方差分析。最后，讨论单向ANOVA的非参数方法。

第13章讨论流行病研究中的设计和分析方法。介绍最重要的研究设计，包括前瞻性研究、病例-对照研究、横断面研究及交叉设计。介绍了混杂变量（即

---

\* 原书把 $t$ 、卡方分布称为是 Sampling distribution，而国内统计学界常把他们称为“样本统计量的分布”。——译者注

该变量与疾病及暴露变量都有关系)的概念,详细讨论了包括 Mantel-Haenszel 检验及多重 logistic 回归的混杂的控制法。随后探讨了流行病学数据分析中目前最受关注的话题,包括 meta-analysis (再分析法——多于一个研究结果的组合);聚集性的二态数据技术(这个技术可以用于分析重复测量数据,比如对相同的人重复测量牙齿);测量误差方法(在收集暴露变量数据时如有实质性的测量误差时可用此法);等效性研究(等效指两种处理方式之间,没有一种比另一种更好)。

第 14 章介绍人-时间数据的分析法。这章中包含有发病率数据的统计方法,及生存分析的几种方法:Kaplan-Meier 生存曲线估计,对数秩检验及 Cox 比例危险率模型。

贯穿于全书,特别是第 13 章,都在讨论研究设计的基础,包括匹配概念,队列研究,病例-对照研究,回顾性研究,前瞻性研究,灵敏度、特异度及筛选检验的预测值。这些设计出现于实际样本的上下文中。此外,第 7、8、10、11、13 及第 14 章包括有不同研究情况下估计样本量的公式。

书末给出统计推理各种方法的一个流程图(749~753 页),它可以作为本书中各种方法相关联系的参考指南。第 7 至第 14 章中每章末尾的某些练习题,我都让学生去查看这些流程图。目的是锻炼学生正确判断本章方法与另外章中方法之间相互关系的能力。

另外,我提供了一个应用索引,按医学专业分组,总结了本书的所有例子及问题。

## 致谢

我感谢 Debra Sheldon, 已故的 Marie Sheehan 及帮助打印书稿的 Harry Taplin, 帮助准备软盘的 Marion Mcphee。我还感谢书稿的审读者,其中有:San Diego 州立大学的 John E. Alcaraz; Pittsburgh 大学的 Stewart J. Anderson; California 大学 Davis 分校的 Christiana Drake; McMaster 大学的 P. D. M. Macdonald; North Carolina 大学 Chapel Hill 分校的 Craig D. Turnbull; California 大学 Berkeley 分校的 Mark J. van der Laan 及 Kansas 大学医学院的 Dennis Wallace。我还要感谢我的同事 Nancy Cook、Robert Glynn、Cathy Berkey 及 Donna Spiegelman, 他们帮助审读了本版的新材料。

另外,我感谢 Carolyn Crockett, Mary Vezilich, Anne Draus, Greg Draus 及 Linda Purrington 提供编辑建议并帮助准备书稿。

我还感谢 Channing 实验室的同事,主要有已故的 Edward Kass、Frank Speizer、Charles Hennekens、已故的 Frank Polk、Ira Tager、Jerome Klein、James Taylor、Stephen Zinner、Scott Weiss、Frank Sacks、Walter Willett、Alvaro Munoz、

Graham Colditz 及 Susan Hankinson 以及我在哈佛医学院的其他同事，主要有 Frederik Mosteller、Eliot Berson、Robert Ackerman、Mark Abelson、Arthur Garvey、Leo chylack、Eugene Braunwald 及 Arthur Dempster，是他们激励了我去写这本书。我还感激 John Hopper 及 Philip Landrigran 为我们的病例研究提供数据。

最后，我感谢 Leslie Miller、Andrea Wagner、Loren Fishman 及 Frank Santopietro，没有他们在医学上的帮助，本版是不可能出版的。

Bernard Rosner



# 目 录

|  |    |
|--|----|
| 第 1 章 概述 .....                         | 1  |
| 参考文献 .....                             | 4  |
| 第 2 章 描述性统计 .....                      | 5  |
| 2.1 绪言 .....                           | 5  |
| 2.2 位置测度 .....                         | 7  |
| 2.2.1 算术均数 .....                       | 7  |
| 2.2.2 中位数 .....                        | 8  |
| 2.2.3 算术平均与中位数的比较 .....                | 10 |
| 2.2.4 众数 .....                         | 11 |
| 2.2.5 几何平均 .....                       | 12 |
| 2.3 算术平均数的某些性质 .....                   | 13 |
| 2.4 离散性测度 .....                        | 14 |
| 2.4.1 极差 .....                         | 14 |
| 2.4.2 分位数 .....                        | 15 |
| 2.4.3 方差与标准差 .....                     | 16 |
| 2.5 方差与标准差的某些性质 .....                  | 18 |
| 2.6 变异系数 .....                         | 19 |
| 2.7 分组数据 .....                         | 21 |
| 2.8 图示法 .....                          | 24 |
| 2.8.1 条形图 .....                        | 24 |
| 2.8.2 直方图 .....                        | 25 |
| 2.8.3 茎叶图 .....                        | 25 |
| 2.8.4 盒形图 .....                        | 28 |
| 2.9 病例研究 1: 儿童的神经及心理机能遭受铅暴露的效应研究 ..... | 30 |
| 2.10 病例研究 2: 中年妇女吸烟对骨无机物密度的效应研究 .....  | 31 |
| 2.11 摘要 .....                          | 33 |
| 练习题 .....                              | 33 |
| 参考文献 .....                             | 41 |
| 第 3 章 概率 .....                         | 42 |
| 3.1 绪言 .....                           | 42 |
| 3.2 概率的定义 .....                        | 42 |
| 3.3 某些有用的概率记号 .....                    | 44 |

|  |            |
|--|------------|
| 3.4 概率的乘法规则 .....                                | 46         |
| 3.5 概率的加法规则 .....                                | 48         |
| 3.6 条件概率 .....                                   | 50         |
| * 3.7 Bayes 规则与筛选检验 .....                        | 53         |
| 3.7.1 ROC 曲线 .....                               | 57         |
| 3.8 患病率与发病率 .....                                | 59         |
| 3.9 摘要 .....                                     | 60         |
| 练习题 .....  | 60         |
| 参考文献 .....                                       | 73         |
| <b>第 4 章 离散概率分布 .....</b>                        | <b>75</b>  |
| 4.1 绪言 .....                                     | 75         |
| 4.2 随机变量 .....                                   | 75         |
| 4.3 离散随机变量的概率质量函数 .....                          | 76         |
| 4.3.1 概率分布与样本分布间的关系 .....                        | 77         |
| 4.4 离散随机变量的期望值 .....                             | 78         |
| 4.5 离散随机变量的方差 .....                              | 79         |
| 4.6 离散随机变量的累加分布函数 .....                          | 81         |
| 4.7 排列与组合 .....                                  | 82         |
| * 4.8 二项分布 .....                                 | 85         |
| 4.8.1 使用二项分布表 .....                              | 87         |
| 4.8.2 使用电子表 .....                                | 89         |
| 4.9 二项分布的期望值 .....                               | 91         |
| 4.10 泊松分布 .....                                  | 92         |
| 4.11 泊松概率的计算 .....                               | 96         |
| 4.11.1 使用泊松表 .....                               | 96         |
| 4.11.2 泊松分布的电子表 .....                            | 96         |
| 4.12 泊松分布的期望值与方差 .....                           | 97         |
| 4.13 泊松分布与二项分布的近似 .....                          | 98         |
| 4.14 摘要 .....                                    | 100        |
| 练习题 .....  | 101        |
| 参考文献 .....                                       | 114        |
| <b>第 5 章 连续概率分布 .....</b>                        | <b>115</b> |
| 5.1 绪言 .....                                     | 115        |
| 5.2 一般概念 .....                                   | 115        |
| 5.3 正态分布 .....                                   | 118        |
| 5.4 标准正态分布的性质 .....                              | 121        |
| 5.4.1 使用正态表 .....                                | 122        |
| 5.5 转换 $N(\mu, \sigma^2)$ 分布到 $N(0, 1)$ 分布 ..... | 126        |

|                                     |     |
|-------------------------------------|-----|
| * 5.6 随机变量的线性组合 .....               | 130 |
| 5.6.1 相依性随机变量 .....                 | 131 |
| 5.7 二项分布的正态近似 .....                 | 134 |
| 5.8 泊松分布的正态近似 .....                 | 140 |
| 5.9 摘要 .....                        | 142 |
| 练习题 .....                           | 142 |
| 参考文献 .....                          | 154 |
| <b>第6章 估计</b> .....                 | 155 |
| 6.1 绪言 .....                        | 155 |
| 6.2 总体与样本的关系 .....                  | 156 |
| 6.3 随机数表 .....                      | 157 |
| 6.4 随机化临床试验 .....                   | 161 |
| 6.4.1 随机化临床试验的设计特性 .....            | 163 |
| 6.5 一个分布中均数的估计 .....                | 165 |
| 6.5.1 点估计 .....                     | 165 |
| 6.5.2 均值的标准误差 .....                 | 168 |
| 6.5.3 中心极限定理 .....                  | 171 |
| 6.5.4 区间估计 .....                    | 173 |
| 6.5.5 $t$ 分布 .....                  | 174 |
| * 6.6 疾病研究:中年妇女吸烟对骨无机物密度的效应研究 ..... | 180 |
| 6.7 方差的估计 .....                     | 180 |
| 6.7.1 点估计 .....                     | 180 |
| 6.7.2 卡方分布 .....                    | 181 |
| 6.7.3 区间估计 .....                    | 183 |
| 6.8 二项分布的估计 .....                   | 185 |
| 6.8.1 点估计 .....                     | 185 |
| 6.8.2 区间估计——正态理论法 .....             | 186 |
| 6.8.3 区间估计——精确法 .....               | 187 |
| 6.9 泊松分布的估计 .....                   | 189 |
| 6.9.1 点估计 .....                     | 189 |
| 6.9.2 区间估计 .....                    | 190 |
| 6.10 单侧置信区间 .....                   | 192 |
| 6.11 摘要 .....                       | 194 |
| 练习题 .....                           | 194 |
| 参考文献 .....                          | 202 |
| <b>第7章 假设检验:单样本推断</b> .....         | 204 |
| 7.1 绪言 .....                        | 204 |
| 7.2 一般概念 .....                      | 204 |

|   |            |
|---|------------|
| 7.3 正态分布均值的单样本检验:单侧备择                               | 207        |
| 7.4 正态分布均值的单样本检验:双侧备择                               | 214        |
| 7.4.1 单样本的 $z$ -检验                                  | 218        |
| 7.5 检验的功效   | 219        |
| 7.5.1 单侧备择  | 219        |
| 7.5.2 双侧备择  | 224        |
| 7.6 样本量的决定  | 226        |
| 7.6.1 单侧备择下求样本量                                     | 226        |
| 7.6.2 双侧备择下求样本量                                     | 229        |
| 7.6.3 基于置信区间宽度的样本量估计                                | 231        |
| 7.7 假设检验与置信区间的关系                                    | 232        |
| 7.8 正态分布中方差的单样本卡方检验( $H_0: \sigma^2 = \sigma_0^2$ ) | 234        |
| 7.9 二项分布的单样本检验                                      | 237        |
| 7.9.1 正态理论法   | 237        |
| 7.9.2 精确方法  | 239        |
| 7.9.3 功效及样本量的估计                                     | 241        |
| 7.10 泊松分布的单样本推断                                     | 242        |
| *7.11 病例研究:中年妇女吸烟对骨无机物密度的效应研究                       | 247        |
| 7.12 摘要   | 248        |
| 练习题   | 249        |
| 参考文献  | 259        |
| <b>第8章 假设检验:两样本的推断</b>                              | <b>260</b> |
| 8.1 绪言  | 260        |
| 8.2 匹配 $t$ 检验                                       | 261        |
| 8.3 两匹配样本均值比较的区间估计                                  | 265        |
| 8.4 等方差的两独立样本均值比较的 $t$ 检验                           | 266        |
| 8.5 两独立样本均值比较的区间估计(等方差情形)                           | 269        |
| 8.6 两方差的相等性检验                                       | 271        |
| 8.6.1 $F$ 分布  | 272        |
| 8.6.2 $F$ 检验  | 273        |
| 8.7 不相同方差下两个独立样本的 $t$ 检验                            | 277        |
| 8.8 病例研究:铅暴露对儿童神经及心理机能的效应研究                         | 282        |
| 8.9 奇异值的处理  | 283        |
| 8.10 均值比较中样本量及功效的估计                                 | 289        |
| 8.10.1 样本量的估计                                       | 289        |
| 8.10.2 功效的估计  | 290        |
| *8.11 纵向研究中样本量的估计                                   | 291        |
| 8.12 摘要   | 294        |

|   |            |
|---|------------|
| 练习题 .....   | 295        |
| 参考文献 .....  | 314        |
| <b>第9章 非参数检验 .....</b>                            | <b>317</b> |
| 9.1 绪言 .....                                      | 317        |
| 9.2 符号检验(匹配数据) .....                              | 318        |
| 9.2.1 正态化理论法 .....                                | 319        |
| 9.2.2 精确方法 .....                                  | 322        |
| 9.3 Wilcoxon 符号-秩检验(匹配数据) .....                   | 323        |
| 9.4 Wilcoxon 秩-和检验 .....                          | 328        |
| 9.5 病例研究:铅暴露对儿童神经及心理机能的效应研究 .....                 | 332        |
| 9.6 摘要 .....                                      | 333        |
| 练习题 .....   | 333        |
| 参考文献 .....  | 339        |
| <b>第10章 假设检验:类型数据 .....</b>                       | <b>340</b> |
| 10.1 绪言 .....                                     | 340        |
| 10.2 二项比例问题的两样本检验 .....                           | 340        |
| 10.2.1 正态理论法 .....                                | 341        |
| 10.2.2 列联表法 .....                                 | 344        |
| 10.2.3 使用列联表法做显著性检验 .....                         | 347        |
| 10.2.4 $2 \times 2$ 列联表中 Yate-修正卡方检验的简单计算公式 ..... | 352        |
| 10.3 Fisher 精确检验 .....                            | 354        |
| 10.3.1 超几何分布 .....                                | 356        |
| 10.4 两样本匹配数据中,二项比例的检验(McNemar's Test) .....       | 359        |
| 10.4.1 正态理论检验 .....                               | 361        |
| 10.4.2 精确检验 .....                                 | 363        |
| 10.5 在两个二项比例的比较中,样本量及功效的估计 .....                  | 366        |
| 10.5.1 独立样本 .....                                 | 366        |
| 10.5.2 配对样本 .....                                 | 368        |
| 10.5.3 临床试验中的样本量及功效 .....                         | 370        |
| 10.6 $R \times C$ 列联表 .....                       | 374        |
| 10.6.1 $R \times C$ 列联表中关联性的检验 .....              | 374        |
| 10.6.2 二项比例中倾向性的卡方检验 .....                        | 377        |
| 10.6.3 Wilcoxon 秩-和检验与倾向性卡方检验的关系 .....            | 381        |
| 10.7 卡方拟合优度检验 .....                               | 383        |
| 10.8 卡帕统计量 .....                                  | 386        |
| 10.9 摘要 .....                                     | 390        |
| 练习题 .....   | 392        |
| 参考文献 .....  | 406        |

|  |     |
|--|-----|
| <b>第 11 章 回归和相关方法</b> .....                      | 408 |
| 11.1 绪言 .....                                    | 408 |
| 11.2 一般概念 .....                                  | 408 |
| 11.3 拟合回归直线——最小二乘法 .....                         | 412 |
| 11.4 回归线中关于参数的推断 .....                           | 415 |
| 11.4.1 简单线性回归的 $F$ 检验 .....                      | 417 |
| 11.4.2 简单线性回归的 $t$ 检验 .....                      | 422 |
| 11.5 线性回归的区间估计 .....                             | 425 |
| 11.5.1 回归参数的区间估计 .....                           | 425 |
| 11.5.2 预测值的区间估计 .....                            | 426 |
| 11.6 回归线的拟合优度 .....                              | 428 |
| * 11.7 相关系数 .....                                | 432 |
| 11.7.1 样本相关系数( $r$ )与总体相关系数( $\rho$ )之间的关系 ..... | 433 |
| 11.7.2 样本回归系数( $b$ )与样本相关系数( $r$ )的关系 .....      | 433 |
| * 11.8 相关系数的统计推断 .....                           | 435 |
| 11.8.1 对相关系数的单样本 $t$ 检验 .....                    | 435 |
| 11.8.2 相关系数的单样本 $z$ 检验 .....                     | 437 |
| 11.8.3 相关系数的区间估计 .....                           | 440 |
| 11.8.4 检验相关系数的样本量估计 .....                        | 441 |
| 11.8.5 相关系数的两样本检验 .....                          | 443 |
| 11.9 多重线性回归 .....                                | 445 |
| 11.9.1 回归方程的估计 .....                             | 445 |
| 11.9.2 假设检验 .....                                | 448 |
| 11.9.3 拟合的优良性准则 .....                            | 453 |
| 11.10 病例研究:铅暴露对儿童神经及心理机能的效应研究 .....              | 464 |
| 11.11 偏相关和多重相关 .....                             | 471 |
| 11.11.1 偏相关 .....                                | 471 |
| 11.11.2 多重相关 .....                               | 472 |
| 11.12 秩相关 .....                                  | 472 |
| 11.13 摘要 .....                                   | 477 |
| 练习题 .....  | 479 |
| 参考文献 .....                                       | 487 |
| <b>第 12 章 多组样本的推断</b> .....                      | 489 |
| 12.1 单向方差分析绪言 .....                              | 489 |
| 12.2 单向方差分析——固定效应模型 .....                        | 489 |
| 12.3 单向方差分析(ANOVA)的假设检验——固定效应模型 .....            | 490 |
| 12.3.1 用 $F$ 检验作组间均数的综合比较 .....                  | 491 |
| 12.4 单向方差分析(ANOVA)中,在指定组之间作比较 .....              | 495 |

|  |            |
|--|------------|
| 12.4.1 指定两组之间作比较的 $t$ 检验 .....                     | 495        |
| 12.4.2 线性约束 .....                                  | 500        |
| 12.4.3 多重比较——Bonferroni 法 .....                    | 502        |
| 12.4.4 线性约束下的多重比较 Scheffé 法 .....                  | 505        |
| * 12.5 病例研究:铅暴露对儿童神经及心理机能的效应研究 .....               | 507        |
| 12.5.1 单向 ANOVA 的应用 .....                          | 507        |
| 12.5.2 单向 ANOVA 与多重回归的关系 .....                     | 510        |
| 12.5.3 单向协方差分析 .....                               | 514        |
| 12.6 双向方差分析 .....                                  | 516        |
| 12.6.1 双向 ANOVA 中的假设检验 .....                       | 517        |
| 12.6.2 双向协方差分析 .....                               | 521        |
| 12.7 Kruskal-Wallis 检验 .....                       | 523        |
| 12.7.1 Kruskal-Wallis 检验中的两两比较 .....               | 526        |
| 12.8 单向 ANOVA——随机效应模型 .....                        | 528        |
| 12.9 组(或类)内相关系数 .....                              | 534        |
| 12.10 摘要 .....                                     | 538        |
| 练习题 .....  | 538        |
| 参考文献 .....   | 550        |
| <b>第 13 章 流行病研究中的设计与分析技术</b> .....                 | <b>551</b> |
| 13.1 绪言 .....                                      | 551        |
| 13.2 研究设计 .....                                    | 551        |
| * 13.3 类型(属性)数据的效应测度 .....                         | 554        |
| 13.3.1 危险率差 .....                                  | 554        |
| 13.3.2 危险(率)比(相对危险度) .....                         | 556        |
| 13.3.3 优势比 .....                                   | 557        |
| 13.3.4 优势比的区间估计 .....                              | 561        |
| 13.4 混杂和分层 .....                                   | 563        |
| 13.4.1 混杂 .....                                    | 563        |
| 13.4.2 标准化 .....                                   | 566        |
| 13.5 分层的类型数据统计推断方法——Mantel-Haenszel 检验 .....       | 568        |
| 13.5.1 分层数据中优势比的估计 .....                           | 572        |
| 13.5.2 效应的修正 .....                                 | 574        |
| 13.5.3 匹配研究中优势比的估计 .....                           | 576        |
| 13.5.4 存在混杂时,患病率有趋势性的检验——Mantel-Extension 检验 ..... | 577        |
| 13.6 分层类型数据中功效及样本量的估计 .....                        | 580        |
| 13.7 多重 logistic 回归 .....                          | 582        |
| 13.7.1 绪言 .....                                    | 582        |
| 13.7.2 一般模型 .....                                  | 583        |

|               |                       |            |
|---------------|-----------------------|------------|
| 13.7.3        | 回归参数的解释               | 584        |
| 13.7.4        | 假设检验                  | 589        |
| 13.7.5        | 多重 logistic 回归中的预测    | 594        |
| 13.7.6        | logistic 回归模型拟合优良性的估价 | 597        |
| * 13.8        | 再分析                   | 602        |
| 13.8.1        | 优势比的齐性检验              | 606        |
| * 13.9        | 等效性研究                 | 608        |
| 13.9.1        | 绪言                    | 608        |
| 13.9.2        | 用置信区间做统计推断            | 608        |
| 13.9.3        | 等效性研究中样本量的估计          | 609        |
| 13.10         | 交叉设计                  | 610        |
| 13.10.1       | 处理效应的估价               | 611        |
| 13.10.2       | 剩余效应的估价               | 615        |
| 13.10.3       | 交叉研究中样本量的估计           | 617        |
| * 13.11       | 聚集性的二态数据              | 618        |
| 13.11.1       | 绪言                    | 618        |
| 13.11.2       | 假设检验                  | 619        |
| 13.11.3       | 聚集性二态数据研究中样本量及功效的估计   | 624        |
| * 13.12       | 测量误差方法                | 627        |
| 13.12.1       | 绪言                    | 627        |
| 13.12.2       | 用金标准暴露变量修正测量误差        | 627        |
| 13.12.3       | 没有金标准暴露变量时测量误差的修正     | 631        |
| 13.13         | 摘要                    | 635        |
|               | 练习题                   | 636        |
|               | 参考文献                  | 645        |
| <b>第 14 章</b> | <b>假设检验:人-时间数据</b>    | <b>648</b> |
| 14.1          | 人-时间数据中效应的测度          | 648        |
| * 14.2        | 单样本发病率数据的统计推断         | 650        |
| 14.2.1        | 大样本检验                 | 650        |
| 14.2.2        | 精确检验                  | 651        |
| 14.2.3        | 发病率的置信区间              | 652        |
| 14.3          | 两样本发病率数据的统计推断         | 653        |
| 14.3.1        | 假设检验——一般性考虑           | 653        |
| 14.3.2        | 正态理论检验                | 654        |
| 14.3.3        | 精确检验                  | 655        |
| 14.3.4        | 率比                    | 658        |
| * 14.4        | 人-时间数据的功效及样本量估计       | 660        |
| 14.4.1        | 功效的估计                 | 660        |



|   |            |
|---|------------|
| 14.4.2 样本量的估计 .....   | 662        |
| 14.5 分层的人-时间数据的统计推断 .....   | 664        |
| 14.5.1 假设检验 .....   | 664        |
| 14.5.2 率比的估计 .....  | 667        |
| 14.5.3 不同层间率比的齐性假设检验 .....  | 670        |
| * 14.6 分层的人-时间数据中功效及样本量的估计 .....  | 671        |
| 14.6.1 样本量的估计 .....   | 671        |
| 14.6.2 功效的估计 .....  | 674        |
| 14.7 发病率数据中趋势性的检验 .....   | 676        |
| 14.8 生存分析的绪言 .....  | 679        |
| 14.9 生存曲线的估计:Kaplan-Meier 估计 .....  | 681        |
| 14.9.1 失访数据的处理 .....  | 683        |
| 14.9.2 生存概率的区间估计 .....  | 685        |
| 14.9.3 危险函数的估计:乘积限方法 .....  | 686        |
| 14.10 对数-秩检验 .....  | 687        |
| 14.11 比例危险率模型 .....   | 693        |
| * 14.12 比例危险率模型中功效及样本量估计 .....  | 699        |
| 14.12.1 功效的估计 .....   | 699        |
| 14.12.2 样本量的估计 .....  | 701        |
| 14.13 摘要 .....  | 703        |
| 练习题 .....   | 704        |
| 参考文献 .....  | 709        |
| <b>附录</b> .....   | <b>710</b> |
| 表 1 精确的二项概率 $Pr(X = k) = \binom{n}{k} p^k q^{n-k}$ .....                          | 710        |
| 表 2 精确的泊松概率 $Pr(X = k) = \frac{e^{-\mu} \mu^k}{k!}$ .....                         | 717        |
| 表 3 正态分布 .....  | 721        |
| 表 4 1000 个随机数字表 .....   | 726        |
| 表 5 $t$ 分布的百分位数 ( $t_{d,u}$ ) <sup>a</sup> .....                                  | 727        |
| 表 6 卡方 ( $\chi^2_{d,u}$ ) 分布的百分位数 <sup>a</sup> .....                              | 728        |
| 表 7a 二项概率精确双侧 $100\% \times (1 - \alpha)$ 置信区间 ( $\alpha = 0.05$ ) .....          | 729        |
| 表 7b 二项概率精确双侧 $100\% \times (1 - \alpha)$ 置信区间 ( $\alpha = 0.01$ ) .....          | 730        |
| 表 8 泊松变量期望值 ( $\mu$ ) 的置信区间 .....   | 731        |
| 表 9 $F$ 分布的百分位数 ( $F_{d_1, d_2, p}$ ) .....                                       | 732        |
| 表 10 检验奇异值的统计量:ESD(极端学生化偏差)的临界值 ( $ESD_{n,1-\alpha}, \alpha = 0.05, 0.01$ ) ..... | 735        |
| 表 11 Wilcoxon 符号-秩检验的双侧临界值 .....  | 736        |
| 表 12 Wilcoxon 秩和检验双侧临界值 .....   | 736        |
| 表 13 Fisher $z$ 变换 .....  | 739        |