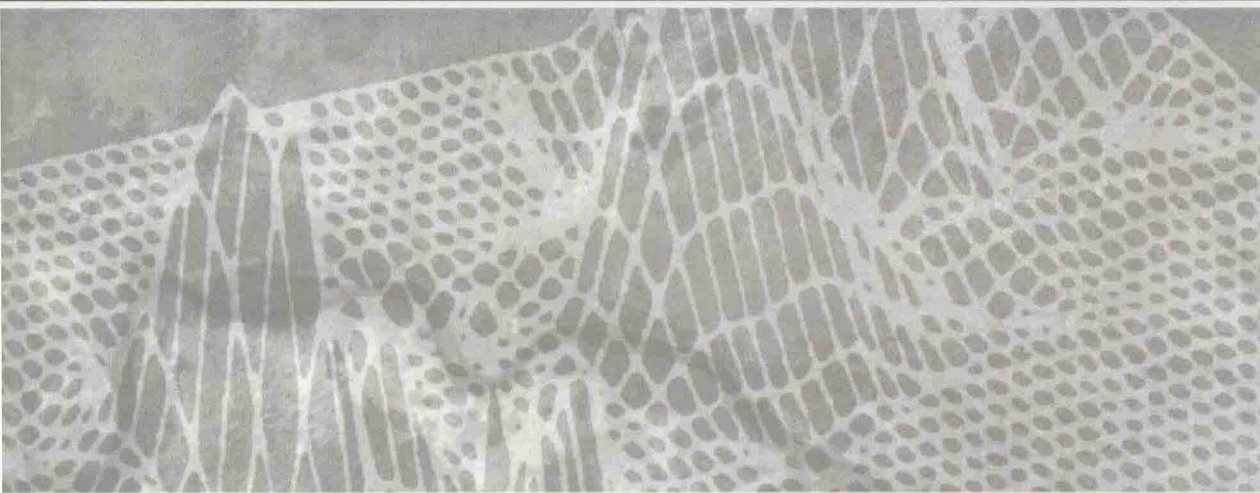


基础地理统计学 (第三版)

Elementary Statistics for Geographers
Third Edition

James E. Burt | Gerald M. Barber | David L. Rigby

(美) 詹姆斯·E. 伯特 (加) 杰拉尔德·M. 巴伯 (美) 戴维·L. 里格比 著
陈江平 邓敏 译



WUHAN UNIVERSITY PRESS

武汉大学出版社

基础地理统计学 (第三版)

Elementary Statistics for Geographers

Third Edition

James E. Burt | Gerald M. Barber | David L. Rigby

(美) 詹姆斯·E. 伯特 (加) 杰拉尔德·M. 巴伯 (美) 戴维·L. 里格比 著
陈江平 邓敏 译



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

基础地理统计学:第三版/(美)杰姆斯·E.伯特,(加)杰拉尔德·M.巴伯,(美)戴维·L.里格比著;陈江平,邓敏译.—武汉:武汉大学出版社,2019.4

书名原文:Elementary Statistics for Geographers, Third Edition
ISBN 978-7-307-11656-6

I.基… II.①杰… ②杰… ③戴… ④陈… ⑤邓… III.地理信息学—统计学—教材 IV.P208.2

中国版本图书馆CIP数据核字(2018)第298065号

Copyright © 2009 by Guilford Press, A Division of Guilford Publications, Inc.

This translation is published by arrangement with Guilford Press. Simplified character Chinese edition
© 2019 by Wuhan University Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the publisher.

本书原名为 *Elementary Statistics for Geographers (Third Edition)*, 作者 James E. Burt, Gerald M. Barber 和 David L. Rigby. 本书中文简体版由 Guilford 出版公司授权武汉大学出版社出版。

版权所有,盗印必究。未经出版者授权,不得以任何形式、任何途径,生产、传播和复制本书的任何部分。

责任编辑:齐翠红

责任校对:汪欣怡

整体设计:韩闻锦

出版发行:武汉大学出版社 (430072 武昌 珞珈山)

(电子邮箱:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:武汉中远印务有限公司

开本:787×1092 1/16 印张:34.5 字数:752千字 插页:2

版次:2019年4月第1版 2019年4月第1次印刷

ISBN 978-7-307-11656-6

定价:118.00元

版权所有,不得翻印;凡购买我社的图书,如有质量问题,请与当地图书销售部门联系调换。

译者前言

这是一本全面介绍地理统计原理与方法的教材。本书由世界知名的三位专家杰姆斯·E. 伯特、杰拉尔德·M. 巴伯和戴维·L. 里格比撰写，是一本在地理空间研究中有统计应用和发展的奠基性教材。它用一种独特的描述方式完美地将统计和时空有机结合。

全书共分四篇。第一篇引言，介绍统计和地理学的关系；第二篇描述性统计，介绍描述性统计的原理、方法及其在地里空间数据中的应用和拓展；第三篇推断统计，介绍推断统计的原理与方法及其在地里空间数据中的应用；第四篇空间与时间模式，介绍了地理数据的空间格局分析与时间序列分析的方法与原理。

本书是地理学、地理国情监测、区域经济学、生态学、环境学、人口学等专业的本科生教材，也可作为研究生教材使用，同时对于相关专业领域的科研工作者也有一定的参考价值，还可以作为地理信息科学相关专业的大学生或研究生的教材。本书也适合各类研究人员参考，包括不具有地理学或统计学背景、但想要快速补充地理空间统计方面知识的工程师，以便在科学研究和应用项目中应用本书的原理与方法。

目前，国内空间统计、空间分析、地学建模类的教材很多，大多从各自专业的角度出发，而如何理解统计方法在地理学中的局限与如何拓展统计方法在地理学中的应用的书籍比较少。为中国学生提供一本有助于快速掌握地理统计的好教材是译者的一个初衷。

基于以上背景，译者利用教学科研之余，着手该教材的翻译，并获得了武汉大学教学改革项目的资助。武汉大学出版社齐翠红和顾素萍女士为本书的编辑和出版做了大量的工作，武汉大学遥感信息工程学院的龚健雅院士、武汉大学测绘遥感信息工程国家重点实验室的舒红教授为本书的翻译提供了很多非常有价值的建议和支持。谨此，向他们表示衷心的感谢。

2019年3月25日

于珞珈山

原著前言

了解我们书籍的读者们会发现本版在原来的基础上做了很多的修订，既包含了大量的新素材，又包含对第二版内容的全面重组。但是本书的目的没有改变：即给大家提供一本易懂的基于代数的教材，并注重解释方法的基本原理和理论基础。如果我们期望学生在自己的研究中能灵活运用统计学方法，或者期望他们采用统计学方法去批判性地评估其他人的工作，那么强调这种理念是很有必要的。此外，当学生了解了第一课所涵盖的方法的基础后，无论他们是在课堂上，还是自学，都能够更好地处理遇到的新概念。我们承认本科生的数学背景通常比较有限，但不认为这是简化课程方法的理由，也不认为学生只需要学习一些已经固有的定量讨论的学科领域而不用参考证明和定量论证。我们常常说如今的学生没有以前的学生善于计算，这种现象也确实可能存在。然而，在我们 20 多年的本科教学中，我们并没有看到学生的能力有所下降，也没有发现他们的智力有所降低。与前几版一样，该书适用于认可这一理念的教师，我们相信他们的学生将从这些理念中受益。

本版为描述性统计部分增加了许多图形化方法的篇幅，用了一整章(第 2 章)来进行讲解。这反映了计算机生成的可视化的结果在统计中的新发展及其日益增长的使用。同样地，学生们也会对可视化学习越来越感兴趣。例如，学生对第 2 章中的“偏度”概念有一个良好的图形认知，他可以利用这种视觉理解更容易地掌握第 3 章中的定量测度。描述部分中出现的另一个新章节是第 4 章：统计关系。这一章介绍了相关和回归的概念和度量。这章的介绍比较“非标准”，因为大多数书籍是在讨论完单变量方法后才讨论它们，而我们发现更早地介绍它们有几个优点。首先，相关和回归是大主题，对于一些学生而言分成两部分学习更好；其次，相关的概念在解释概率论的某些方面，例如独立性、条件概率和联合概率时十分有用；最后，在已经具有统计相关思想的情况下，卡方检验等非参数检验将更容易讨论。当然，那些喜欢在课程的一部分讲授相关和回归的教师可以将第 4 章推迟，并将其作为第 12 章和第 13 章的一部分进行讲述。

推论统计部分也进行了大量修订。我们将基本概率理论与随机变量的处理相结合，以便在单一章节(第 5 章：随机变量与概率分布)中建立更加简化的概述。“计算机密集型方法”一章已不复存在，其大部分内容纳入到“非参数方法”一章。由于自举法(boot-

strapping)和相关技术已成为主流,把那些方法与其他的非参数方法放到一起会更合适。第11章“方差分析”是一个新章节,包含单因素和双因素的设计。另外还有第13章“拓展的回归分析”,它可以处理回归的诊断、变形以及一些更先进的回归模型(包括多元回归)。

最后一部分“空间与时间模式”包含第二版中“时间序列分析”一章的修订版本,以及全新的第14章“空间格局与关系”。后者是空间分析的概述,涵盖点模式(特别是最近邻分析)、空间自相关(变异函数、连接数、Moran's I、LISA 和 G-统计量)以及空间回归(包括地理加权回归的介绍)。

此外,还有许多小的改动不在此一一列出。我们更加注重样例,并且通常为它们配图。我们在整本书中提到的数据集都可以在本书附带的网站上找到。在网站上,读者还可以找到大部分章节末习题的答案。有关在线资源,请参阅 www.guilford.com/pr/burt。

前面已经说过,在定量方法的介绍中,新版本坚持以前版本的理念,强调解释而不是简单描述。在其他方面也没有改变,该版本保留了时间序列的内容,它在这种类型的书中很少被涉及。时间序列数据在地理学的各分支中都很常见,因此地理学者至少需要配备数种时间数据分析的工具。此外当学生接触线性回归时,他们也将有能力了解时间序列建模的基础知识。换句话说,处理时间序列的能力可以以较少的额外成本获得。由于时间序列是如此的普遍,无论地理学者在学科中的正式培训如何,他们都可能会在工作中涉及时间数据。我们认为,即使是简单的操作,如运行程序,也必须了解程序含义。因为大多数学生不会完成全部的时间序列课程,所以至少在介绍文本中涉及时间序列是至关重要的。此外,我们还从教学中获得了关于此部分材料的一些积极反馈。

本版坚持我们在第二版中的做法,不将这本书链接到任何特定的软件包。我们相信大多数教师会使用软件来讲授这本教材,但没有哪种软件包能够超过所有的而成为首选。将本书装载到一个特定的软件包中可能会照顾到使用同种软件包的用户,但相对地则不利于使用其他软件包的人。此外,由于统计软件变得越来越易用,学生对教学方式要求也逐渐变少。我们希望这本书保持现状。我们已经发现,即使是输出格式的细微变化,也可能使那些被引导去寻找特定值表或软件包中特定术语的学生感到困惑。

最后,为了配合版本的发展趋势,这本原本很长的书变得甚至更长了。除非是在长达一年的课程中使用,否则教师必须非常有选择性地分配讲授内容。考虑到这一点,我们试图使这些章节尽可能相互独立。除了关于概率和抽样理论的章节外,使用挑拣章节的方法就能够合理分配。例如我们从经验中知道,一些教师跳过了整个“非参数方法”章节,而并没有产生下游效应。其他跳过章节内的各小节的教师也是如此。如果有些学生抱怨跳过了太多章节,他们中的大多数已经认识到这本书涵盖的比课堂讲授的要多。之后当他们在阅读或研究项目中遇到不熟悉的方法时,他们可以重回这本记着他们熟悉而奇怪注释的书,并且还能在课程介绍过的内容中了解到新的技术。当我们

回过来看自己的书架时，被证明多年来对我们最有用的书正是这种类型的书籍。我们不会擅自认为我们的工作具有类似的持久效用，但是我们相信涉及面过多要优于过少。

非常感谢帮助我们编写这本书的人们。特别感谢加州大学洛杉矶分校、皇后大学和威斯康星大学麦迪逊分校的学生和助教，他们告诉了我们哪些方法有用而哪些没有用。感谢匿名评审小组对原稿版本的评论，他们大大地提高了本书的质量。我们也非常感谢吉尔福德新闻社中参与该项目的每个人的努力工作。特别是本书的编辑 Kristal Hawkins，充满了耐心与鼓励，编辑 William Meyer 还特别提到了他对项目的打印和对数字组件的特别关注。最重要的是，要感谢我们的家属如此乐意于不辞辛劳地关注我们。对所有这些人，我们唯有奉献这本书。

目 录

- I. 引言 1
- 1 统计和地理 3
 - 1.1 统计分析和地理学 8
 - 1.2 数据 13
 - 1.3 测量评估 24
 - 1.4 数据和信息 26
 - 1.5 总结 27
- II. 描述性统计 31
- 2 数据的显示和解读 33
 - 2.1 定性变量分布的显示及解读 34
 - 2.2 定量变量分布的显示与解读 39
 - 2.3 显示和解读时间序列数据 61
 - 2.4 空间数据的显示和解读 66
 - 2.5 总结 76
- 3 用统计来描述数据 80
 - 3.1 集中趋势的度量方法 80
 - 3.2 离差的度量方法 92
 - 3.3 高阶矩及量测变量分布特性的其他数值方法 99
 - 3.4 使用时序数据的描述性统计 100
 - 3.5 空间数据的描述性统计 106
 - 3.6 总结 126
 - 附录 3a $\sum$ 符号 127
 - 附录 3b 确定欧氏中位数的迭代算法 129
- 4 统计关系 135
 - 4.1 相关性和独立性 136

4.2 通过图和表查看关系	137
4.3 相关性介绍	144
4.4 回归	149
4.5 时间自相关	163
4.6 总结	166
附录 4a 线的基本几何性质	167
附录 4b 利用初等微积分求最小二乘解	169
III. 推断统计	173
5 随机变量与概率分布	175
5.1 基本概率论	175
5.2 随机变量的概念	182
5.3 离散型概率分布模型	190
5.4 连续型概率分布模型	201
5.5 二维随机变量	204
5.6 总结	211
附录 5a 概率计算的计数规则	211
附录 5b 连续型随机变量的期望值和方差	213
6 抽样	217
6.1 我们为什么抽样	218
6.2 抽样的步骤	219
6.3 抽样的类型	222
6.4 概率随机抽样和相关设计	224
6.5 抽样分布	231
6.6 地理抽样	238
6.7 小结	245
7 点估计和区间估计	248
7.1 统计估计方法	249
7.2 点估计	253
7.3 区间估计	255
7.4 样本容量的确定	264
7.5 总结	267
8 单样本假设检验	270
8.1 古典假设检验的主要步骤	270
8.2 假设检验中的 PROB-VALUE 方法	280
8.3 关于总体均值 μ 和 π 的假设检验	283

8.4	假设检验与置信区间估计的联系	290
8.5	统计显著性 vs 实际显著性	291
8.6	总结	292
9	双样本假设检验	296
9.1	均值差	296
9.2	成对观测的均值差	303
9.3	比例差	306
9.4	方差齐性	308
9.5	总结	310
10	非参数检验	314
10.1	参数检验与非参数检验的比较	315
10.2	单样本与双样本检验	317
10.3	多样本 Kruskal-Wallis(克鲁斯卡沃利斯)检验	327
10.4	拟合优度检验	331
10.5	列联表	338
10.6	估计一个概率分布: 核估计	340
10.7	自举法(Bootstrapping, Bootstrap 方法)	348
10.8	总结	355
11	方差分析	360
11.1	单因素完全随机设计	362
11.2	双因素完全随机设计	371
11.3	Scheffé 成对比较检验	377
11.4	方差分析的假设	379
11.5	总结	380
	附录 11a 从公式(11-10)推导公式(11-11)	380
12	线性回归的推理	384
12.1	回归分析中的步骤概述	384
12.2	简单线性回归模型的假设	387
12.3	回归分析中的推论	396
12.4	线性回归模型的图形诊断	405
12.5	总结	411
13	拓展的回归分析	414
13.1	多元回归分析	414
13.2	变量变换和回归函数的形状	427
13.3	验证回归模型	436
13.4	总结	439

IV. 空间与时间模式 443

14 空间格局与关系 445

 14.1 点格局分析 445

 14.2 空间自相关 453

 14.3 空间关联的局部指标 467

 14.4 空间自相关数据回归模型 473

 14.5 地理加权回归 476

 14.6 总结 477

15 时序分析 482

 15.1 时序过程 482

 15.2 随机过程的特性 484

 15.3 随机过程的类型 488

 15.4 去趋势：转换到平稳 491

 15.5 模型检验 492

 15.6 模型拟合 496

 15.7 时序模型，移动均值，过滤器 501

 15.8 频率变换法 502

 15.9 滤波器的设计 508

 15.10 总结 514

附录 统计表 517

关于作者 540

I. 引言

统计和地理

我们大多数人首次接触到概率和统计是通过广播、电视、报纸或杂志。我们可能会看到或听到关于政治民意调查或者癌症、心脏疾病治疗的最新进展研究和调查报告。如果我们自己思考一下，可能会注意到，统计几乎被运用在人类活动的所有领域。例如，许多体育组织对观众进行统计；许多公司为了完成用于运行目前的生产和配送系统的监控，需要收集大量数据。或许世界上最大的数据收集工作，就是由各国政府定期完成的，如全国人口普查。

“统计”一词还有另外一个更专业化的意义。它是指收集、展示和分析数据的方法。这个方法是许多学科领域调查的基础，如教育、物理学、工程学、医学、生物科学和包含地理学在内的社会科学等。即使是在传统的非定量统计的人文学科中，这种方法的使用也一直在增加。

定义(统计) 统计是用于研究如何通过图形和数值的形式，收集、整理、概括数据，并对数据进行分析，最终得出结论的一种方法。

许多学生接触统计学，是为了理解和解释自己感兴趣领域的研究。因此，他们必须有这些领域的专业基础知识，如过程、符号和专业词汇。

无论哪一门学科使用到的统计方法，都是从数据收集开始分析的。对数据分析的目的通常出于以下几点：

(1) 为了帮助总结某些调查的结果，例如，研究老年人或残疾人士的旅游行为，或者是评估植树造林的木材要求。

(2) 在研究过程中，为了更好地理解某一现象，用来帮助找到某种一般规律，或者是找到理论上的验证，例如，验证城市地租理论。

(3) 对某些变量做预测，例如，短期利率、选民的行为或者房价。

(4) 为了评估某些项目的性能，例如饮食形式、创新医疗、教育计划或改革。

(5) 用来从很多种选择中选一种实施方案，或规划某些系统，例如，学校的选址。

统计方法的原理可以在各种情况下使用，说明其用途广泛。

统计方法可以分为两部分：描述性统计和推断统计。描述性统计是用来组织和概括数据的。描述性统计的目的是在一些数据集中用少量的概要的知识来替换那些海量的大数据集。每当发生这样的替换时，不可避免地存在一些信息损失。用较小的数据集来保存大数据集中的所有信息是不可能的。描述性统计的主要目标之一是尽量减少这种信息损失的影响。了解哪些统计措施应该在特定情况下用作概要索引是描述性统计的另一个重要统计目标。如果我们理解了描述性统计的推导和使用，并意识到它的局限性，就可以帮助我们避免错误结果的传播。对统计的不信任源于对一些统计方法的误用以及相关统计结果被不适当地运用或解释。正如摄影师可以使用一个透镜来扭曲场景，统计学家可以通过选择的概述数据的统计方法，改变数据集的信息。了解描述性统计能告诉我们什么，以及不能告诉我们什么，是统计分析的一个关键问题。

统计方法的第二个主要部分是推断统计。描述性统计一般与概率论联系在一起，以便研究者可以将一些个人的研究结果推广到一些较大的群体。为了解释这个过程，有必要介绍一些简单的定义。在研究人口的过程中，人、地区或对象的集合都是研究者感兴趣的。

定义(统计总体) 统计总体是一组特定研究元素的全部的集合(如对象、个人、地区、街道、河流等)。

例如，如果一个地理学家正在研究某个区域内的农场实践，相关的统计总体包括在某一特定日期或特定的时间段内该地区的所有农场。下一个例子，对一个城市的选民行为统计的研究，统计总体包括所有的潜在选民，这些人通常包含在一个合格的选民名单之中。

在许多情况下，我们所考虑的统计总体是有限的；也就是说，在总体中的每个元素可以被列出。在某市或者县中，合格选民名单和取样调查的评估问卷，这些都是统计总体有限的例子。在其他时候，可能要对总体进行假设，例如，钢铁制造商想测试一批产品的质量，他可能会从几个星期内生产的产品中选 100 个铸件来测试。在这个例子中，统计总体是制造商用该设备来生产的未来产品。当然，这个统计总体不存在，并且可以具有无限多的元素。统计分析与有限的假设的总体是相关的。

一般，我们感兴趣的是总体中的一个或多个特征。

定义(总体特征) 总体特征是对总体中一个元素进行统计的任何可测量属性。

河流地貌学家要研究一条河流分水岭，可能会研究这条河流不同的可测量属性，如

水流速度、流量、输沙量等诸多属性数据可能会在实地研究期间收集。由于总体特征通常取决于总体的不同属性的值，一个属性通常被称为一个变量。由于总体统计特征取不同的值，使得统计推断的过程必须成为事实。如果总体特征并不随总体的变化而变化，从这个角度来说，研究者将不太会对这个特征感兴趣。

定义(变量) 变量是总体的一个特性，即构成总体的取不同值的元素。

关于总体信息可以用两种方式来收集。第一种方法是要确定个体和总体感兴趣的变量的值。这在人口普查或人口详查中比较常见。很显然，这只是对有限的人群一个可行的选择。因为总体数量太多，这是非常困难的，有人甚至会说是不可能的。实际上，在一个大人口的国家每十年一次的人口普查中一次普查所有的个体，这是不可能的，但如果详查过程是精心设计的，错误可以被降低到最小。

定义(总体普查) 总体普查是总体的所有元素相关特征的完整列表。

获得关于一个总体信息的第二种方法是通过样本。样本是一个简单的总体的子集，因此在采样中，我们得到的值是所选中的样本的值。

定义(抽样) 抽样是总体的子集，用于对整体的某些特性进行推断。

出于时间和成本考虑，更为方便的是用样本而不是列举整个总体。当然，采样有一个明显的劣势。把我们的研究对象限制在总体的一小部分，这样就不可能像使用完整的总体一样，准确地得到总体的特征，错误的风险也将会增加。

定义(抽样误差) 抽样误差是总体特性的实际值和从样品推断的值之间的差。

为了说明抽样误差，考虑某一年在一个给定的大都市区房屋的平均销售价格的总体特征。如果每家被检查，可以发现，平均售价为 15 万美元。但是，如果每月只抽查 25 家，样品中的 300 家(12 个月×25 家)的平均销售价格，可能是 12 万美元。所有其他的条件都是一样的，我们可以说， $\$150\,000 - \$120\,000 = \$30\,000$ 是由于抽样误差产生的。

这里说的“所有其他的条件都是一样的”是什么意思呢？\$30 000 的误差，部分原因

可能是采样之外的其他因素。也许在一户中，售价样品被错误地标识为 \$252 000，而不是 \$152 000。许多这种类型的错误发生在大型数据集中。从个人访谈或问卷调查中获得的信息可能包含受访者的事实错误，比如缺乏回忆、无知或仅仅是受访者不太坦诚。

定义(非抽样或数据采集错误) 在采集、记录和编辑统计数据时出现的错误称为非抽样或数据采集错误。

为了使样本和总体之间的差别或错误可以单独归因于抽样误差，最小化非抽样误差是非常重要的。验证检查、精心编辑、仪器校准等方法是用以减少由非抽样误差导致总误差显著增加的措施，否则会扭曲随后的统计推断。

概率论是研究样本和总体之间的联系理论。关于总体的推断是基于样本中的信息。这些推断的质量取决于样本是否能够很好地体现或代表总体的信息。不幸的是，不全面普查，没有办法知道一个样本是否很好地体现了总体的性质。因此，我们选择随机抽样，而不是选择具有代表性的样本。

定义(代表性样本) 代表性样本是样本特征与总体特征非常一致的样本。

定义(随机抽样) 随机抽样是在总体中的每一个个体都有相同的机会或者概率被选入样本中的一种抽样方法。

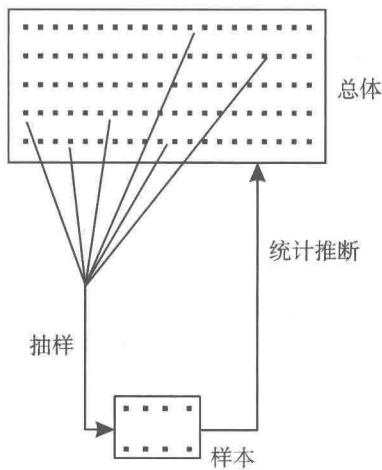


图 1-1 统计推断的过程

基于随机抽样的统计推断能确保调查结果的公正。选到一个非常不具代表性的随机样本是有可能的，但如果样本足够大，选到这种样本的概率是非常小的。事实上，因为样本是被随机选择的，我们总能确定从样品获得错误推论的概率。这就是为什么统计学家总是作出概率判断，从来不作确定的判断。推断总是限定在某种程度上，随机抽样误差可能导致错误的判断。

统计推断的过程如图 1-1 所示。总体中的成员或单元，是在抽样过程中被选中的。这些单元合在一起组成了样本。有关总体的推论，就是从这些样本中来的。简而言之，