数据分析与模拟丛书

Song S. Qian 著
曾思育 译

Environmental and Ecological Statistics with R

环境与生态统计

——R 语言的应用

高等教育出版社
HIGHER EDUCATION PRESS

数据分析与模拟丛书

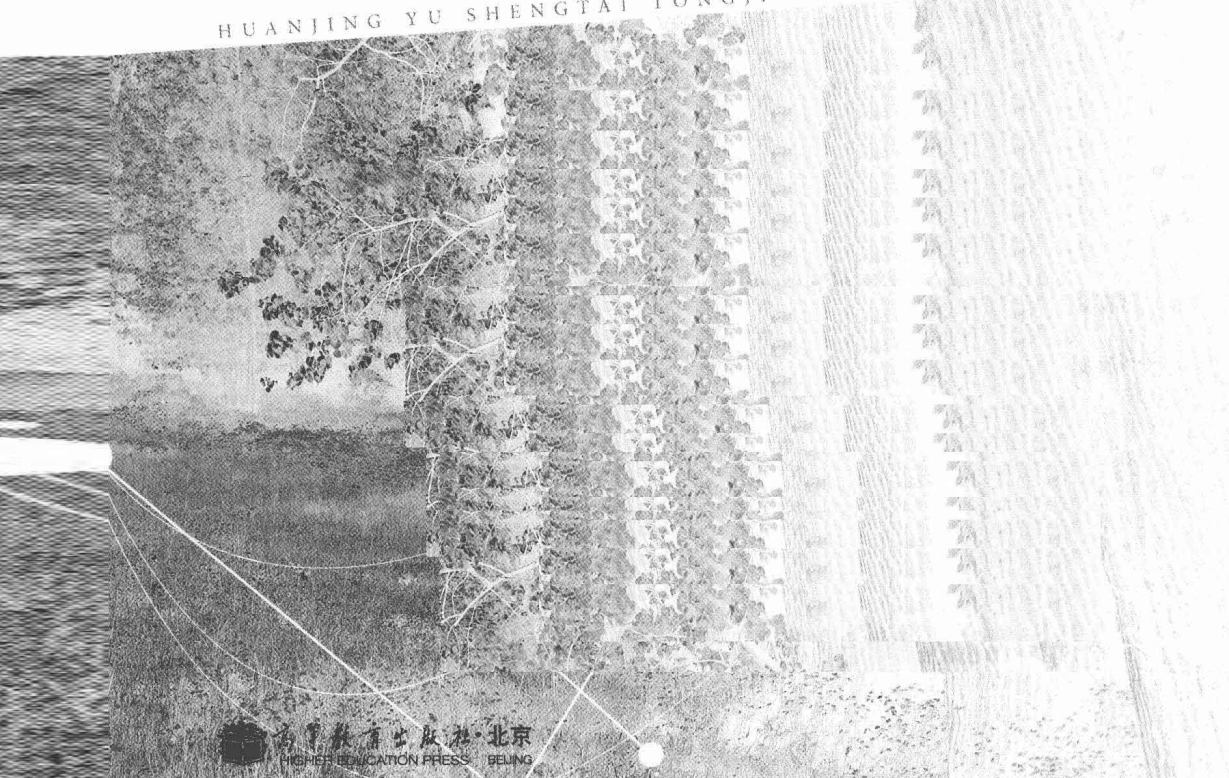
Song S. Qian 著
曾思育 译

Environmental and Ecological Statistics with R

环境与生态统计

——R 语言的应用

HUANJING YU SHENGTAI TONGJI — R YUYAN DE YINGYONG



高等教育出版社·北京
HIGHER EDUCATION PRESS BEIJING

图字：01-2010-5363 号

图书在版编目（CIP）数据

环境与生态统计：R 语言的应用/钱松著；曾思育译. —北京：高等教育出版社，2011.7

书名原文：Environmental and Ecological Statistics with R

ISBN 978-7-04-031893-7

Ⅰ. ①环… Ⅱ. ①钱… ②曾… Ⅲ. ①环境生态学：环境统计学-统计模型 Ⅳ. ①X171.1 ②X11

中国版本图书馆 CIP 数据核字（2011）第 121828 号

策划编辑 陈正雄	责任编辑 柳丽丽	封面设计 张 楠	版式设计 马敬茹
插图绘制 尹 莉	责任校对 杨凤玲	责任印制 张福涛	

出版发行	高等教育出版社	网 址	http://www.hep.edu.cn
社 址	北京市西城区德外大街 4 号		http://www.hep.com.cn
邮政编码	100120	网上订购	http://www.landraco.com
印 刷	北京天来印务有限公司		http://www.landraco.com.cn
开 本	787 mm×1092 mm 1/16		
印 张	25.25		
字 数	470 千字	版 次	2011 年 7 月第 1 版
购书热线	010-58581118	印 次	2011 年 7 月第 1 次印刷
咨询电话	400-810-0598	定 价	49.00 元

物 料 号 31893-00

译 者 序

2010 年春节过后，有幸拜读了美国杜克大学钱松（Song S. Qian）教授的专著《Environmental and Ecological Statistics with R》。后来受高等教育出版社的委托，我开始着手翻译这本书。之所以接手这项工作，是因为原著对我有两方面的吸引力，希望能亲自将它翻译出版，让更多的人受益。一方面，因为循序渐进、深入浅出的原则始终贯穿在各个章节中，所以原著能够把原本让人人都发憷的统计学方法讲得有声有色，非常有利于读者的学习和掌握。原著从基本概念讲起，对长期以来很容易被大家混淆和误解的一些概念则花费了更多笔墨，让读者知其然又知其所以然。在介绍每种方法时所选择的案例都非常贴近环境与生态科学领域的研究实践，且具有足够的典型性，容易引起读者兴趣，在不知不觉中得到提高。我现在正给清华大学的本科生讲授“环境数据处理与数学模型”课程，涉及的不少统计分析方法都可以在书中找到极好的例子，完全可以把它用作教学参考书。另一方面，每讲到一种模型，原著都配套描述了如何用 R 语言来实现。R 语言的应用可以让我们摆脱昂贵的商业统计软件，为相关的教学和科研工作提供更多的灵活性。利用 R 语言来完成环境与生态科学研究中的数据分析和统计建模任务，是该书非常突出的特色。对 R 语言的讲解详细而实用，使之成为国内目前不可多得的资料。

正是这样一本极具特色和优势的好原著，激励我用了大致一年的时间来完成艰苦的翻译和校对工作，为的是能给读者奉献一本好译著。不过，由于时间和水平有限，翻译过程中难免有疏漏，可能还存在这样或那样的问题，敬请读者批评指正！

曾思育

2011 年 5 月 16 日

于清华园

前　言

统计学是全世界高等教育机构中几乎所有环境和生态学研究院系和培养项目都开设的课程之一。统计学还常被认为是最不讨人喜欢、教学效果差的科目，尤其是对于数学/统计学领域之外的学生更是如此。学习统计学过程中普遍存在的问题是，统计学常被认定是数学的一个分支领域。因此，我们的期望是学到一套规则并能够将统计学应用到我们的工作中。但是，应用统计学并不是数学。本书致力于在典型的应用统计学和环境与生态学领域对统计学的需求之间架起一座桥梁，重点则放在统计学思考过程中的归纳特征上。内容上避开了很多数学和理论背景，通过案例进行概念的介绍和方法的阐释。如同R. A. Fisher将统计学引入应用科学那样，本书将统计学当做一种推动科学思考的工具加以介绍。

本书所采用的方法遵循了R. A. Fisher关于统计建模的一般步骤，即模型定义、参数估计以及模型评估。这些步骤与研究人员在科研项目中所采取的步骤是相似的。但是，正如很多人讨论的那样，统计学往往不是科学和工程领域的学生们所喜欢的话题（Berthouex和Brown，1994），生态学家们也常常会在这上面犯错误（Peters，1991）。这是由于在典型的应用统计学课程/教材和典型的科学问题之间缺乏联系。求解一个科学问题时，我们先从潜在机理的假设开始，以之作为收集数据的依据。所提出的假设提供了用公式表达模型的基础，公式中则常常带有未知参数。实验和其他数据收集工作则为估计这些未知参数提供数据。一旦估计出这些参数的值，科研人员就可以通过比较模型的预测值和新的观测值来评价模型。在对科学问题求解过程的简单概括中，第一步（形成假设）往往是最困难的部分，需要研究人员既要有经验又要有创造性。因为一个错误的模型绝不可能把我们带向成功，所以模型/假设的公式表达是这一过程中最重要的步骤。在应用统计学中，正如R. A. Fisher所描述的那样，我们所遵循的典型步骤与求解科学问题的步骤是相似的。对于一个具体问题，我们首先必须考察数据，然后提出统计模型来描述我们感兴趣的变量的分布。统计学模型的参数化靠的是那些需要利用数据进行估值的未知参数。当完成参数估值后，我们必须通过检验估计出的参数值的抽样分布来评估模型的不确定性。然而，对于研究人员而言，因为困难在于从科学假设到统计学模型的转换，所以科学问题求解和统计学模型开发过程中的相似性并没有解决统计学学

习中的困难。典型的应用统计学课程/教材将这一科目当做不同类型统计学模型的方法汇总来讲授，或多或少地忽视了模型公式表达中的问题。因为模型的公式表达必然是一个科学问题，所以这种处理是不可避免的。应用统计学教材或者课程集中于讨论参数估值和模型评估中的问题。不同类型的模型往往需要不同的数学求解。这样对待统计学通常会导致大家错误地去认识什么是统计学和我们为什么要学习统计学。

本书的灵感来自于统计学思考和科学方法之间潜在的联系。本书以统计学模型为基础来组织。但是，通过贯穿全书的案例来讨论每种类型的统计学模型，其中一些例子覆盖了几种模型。这些案例的重点是模型的公式化，背后的数学/统计学理论则大部分被略去了，代之以介绍如何用 R 语言来实现这些模型。本书的基础是我在杜克大学环境学院积累的教学资料。本书划分为 3 个单元。

第 1 章到第 5 章曾被用于研究生水平的应用数据分析课程，可以作为高级统计建模的预备知识来阅读。这些章节的目的是打好基础，以便读者能够开展简单的数据分析工作，如探索性数据分析和拟合线性回归模型等。

第 6 章到第 8 章曾被用于统计建模的后续课程。本单元中的这 3 章之间基本上是相互独立的，可以分别阅读。第 8 章中的 3 个主题（第 8.1—8.4 节、第 8.5 节和第 8.6 节）也是一样的情况。

第 9 章和第 10 章曾被用于博士水平的独立研究课程。第 9 章讨论了如何使用模拟手段进行模型检验，为开发出的模型提供了评估鉴定的工具。模拟方法在参数估值和不确定性评估中是会普遍使用的。虽然在文献中对使用模拟方法来检验模型的讨论并不多，但它是模型开发与评估的重要方面。第 10 章讨论了多层回归模型的应用，这是一类可以对环境和生态学数据分析产生广泛影响的模型。

本书中使用的数据和 R 脚本可以在线获取：http://www.duke.edu/~song/eeswithr.htm。

很多人为本书的撰写提供了帮助。Kenneth H. Reckhow、Curtis J. Richardson 和 Michael Lavine 是我的导师和长期合作者。本书反映了他们对我走进环境与生态统计学领域的影响。与 Yandong Pan 的合作加强了我对生态学问题和生态学问题求解过程的认识。Craig A. Stow 不断地给我提供有意思的想法和论文，非常感谢他在分析鱼体内 PCB 数据时所做的工作。Olli Malve、George B. Arhonditsis 和 Andrew D. Gronewold 花了无数小时帮我理清思路和概念。Thomas F. Cuffney 和 Gerard McMahon 给了我 EUSE 的案例，并花了大量时间与我讨论第 10 章中的案例。沈泽昊于 2007 年夏天在北京大学接待了我的访问，并提供了很多有趣的案例。Richard L. Smith 阅读了本书的草稿并提出了批

评意见，从而大大改善了本书的表达，对一些关键概念的讨论也使这些概念变得更为清晰。Meg Mobley、Ibrahim Alameddine、Itai Shelem、Kristen Marine、Emily Sharp、Erin Gray 和 Wyatt Hartman 发现了多处错误，并提出了改进建议。

Song S. Qian
于美国北卡罗来纳州 Durham
2009 年 3 月

目　录

第Ⅲ部分 高级统计建模

表 清 单

图 清 单