

# 试验应用统计

## 设计、创新和发现 (原书第2版)

*Statistics for Experimenters Design, Innovation, and Discovery*

(Second Edition)



George E. P. Box  
(美) J. Stuart Hunter 著  
William G. Hunter

张润楚 刘民千 杨建峰 刘玉坤 李鹏 译



机械工业出版社  
China Machine Press

统计学精品译丛

# 试验应用统计

设计、创新和发现 (原书第2版)

*Statistics for Experimenters Design, Innovation, and Discovery*

(Second Edition)

George E. P. Box  
(美) J. Stuart Hunter 著  
William G. Hunter

张润楚 刘民千 杨建峰 刘玉坤 李鹏 译



机械工业出版社  
China Machine Press

本书从试验工作者的角度阐述了统计方法在试验设计中的应用, 强调科学地利用统计工作从试验数据中获取最大的信息。内容主要包括: 基础知识、比较两总体、两水平因析设计、部分因析设计、因析设计及数据变换、变差的多种来源、最小二乘与试验设计的必要性、响应曲面的某些应用等。

本书内容丰富, 从实际问题出发, 分析各种方法的利弊, 然后采用最佳统计方法解决问题。本书适合作为理工科各专业本科生、研究生的统计学教材, 也可作为相关领域研究人员的参考读物。

George E. P. Box, J. Stuart Hunter, William G. Hunter: Statistics for Experimenters: Design, Innovation, and Discovery, Second Edition (ISBN 978-0471-71813-0).

Authorized translation from the English language edition published by John Wiley & Sons, Inc.

Copyright © 2005 by John Wiley & Sons, Inc.

All rights reserved.

本书中文简体字版由约翰·威利父子公司授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

版权所有, 侵权必究

本书法律顾问 北京市展达律师事务所

本书版权登记号: 图字: 01-2005-5611

图书在版编目(CIP)数据

试验应用统计: 设计、创新和发现/(美)伯克斯(Box, G. E. P.)等著; 张润楚等译.  
—北京: 机械工业出版社, 2009.10

(统计学精品译丛)

书名原文: Statistics for Experimenters: Design, Innovation, and Discovery, Second Edition  
ISBN 978-7-111-27258-8

I. 试… II. ①伯…②张… III. 统计分析(数学)-应用-试验设计(数学) IV. 0212

中国版本图书馆CIP数据核字(2009)第082628号

机械工业出版社(北京市西城区百万庄大街22号 邮政编码 100037)

责任编辑: 王春华

北京京北印刷有限公司印刷

2010年1月第1版第1次印刷

186mm×240mm·25.25印张

标准书号: ISBN 978-7-111-27258-8

定价: 65.00元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991; 88361066

购书热线: (010) 68326294; 88379649; 68995259

投稿热线: (010) 88379604

读者信箱: hzsj@hzbook.com

## 译者序

现在越来越多的人意识到统计学是人类认知世界的三门重要公共基础科学之一，即除了哲学、数学（这里主要指纯数学）之外，就是统计学。哲学是通过对世界（客观或主观）总的看法和思维逻辑方式的研究来认知世界的，数学是通过对世界（客观或主观想像）的空间和形态（包括存在于现实或思维可能想像的数量、符号、图形等）的逻辑关系的研究来认知世界的，而统计学则是通过对世界（客观或主观）的观察，即通过观察世界发生的现象而得到各种形态的信息、数据资料（包括被动得到的或者主动设计收集而得到的）来认知世界的，包括研究如何观察和收集信息、数据以及对得到的数据如何进行分析的方法。这三门公共基础科学既各自具有一定的独立性，又相互联系和相互支撑，哲学和数学更具有综合指导性，统计学要用到前两者的思维方法和知识，但又为前两者的研究提供参考或依据，而且它们的理论和方法都要通过时间和实践来检验。这里实践的含义可能各自有所不同，例如，哲学的理论要得到各类其他科学的验证（包括实际的检验），数学要经得起思维逻辑严密性推敲或者在其他科学领域应用的有效性来检验，而统计学则更多地要得到实际的直接检验。要提高人们的科学素养和认知能力，这三者看来都是不可缺的。

作为认知和思维方法，试验设计和分析是统计学中最重要的分支之一。从历史上讲，试验设计为现代统计学的建立打下了基础。现代统计学最重要奠基人之一的伟大统计学家 R. A. Fisher，早在 20 世纪 20 年代就从研究农业试验设计开始研究统计学，为建立现代统计学进一步奠定基础。现代试验设计也就从此诞生，后来得到迅速发展，成为统计学中开发最早、影响最大的分支之一。至今，已形成广泛的理论和应用体系，并已经广泛地应用到涉及自然科学和技术的工业、农业、生物、医学、卫生、物理、化学、工程、环保、科学计算、仿真技术、航空航天等领域以及人文社会经济科学的各个分支，即几乎所有科学领域。可以说，凡是要通过观察和进行科学实验来了解某个客体、现象或过程并进行统计推断的地方，都要用到试验设计和试验数据分析。尤其是在市场激烈竞争的今天，产品质量被视为企业的生命，试验设计便成为产品质量工程的重要组成部分而受到特别重视，应用十分广阔。

由 G. E. P. Box、J. S. Hunter 和 W. G. Hunter 著的本书（*Statistics for Experimenters*）是为试验者写的统计学教材，介绍了针对试验设计用到的统计学思想和方法，同时生动、概括地描述了什么是统计学以及人们如何用统计学认知世界的思维过程和方法，因此，无论是对试验者还是对一般想了解统计学的人来说，本书都是一本不可多得的好书。无疑，本书自 1978 年出版以来，对试验设计的发展做出了划时代的贡献。此外，它对促进统计学的发展，包括对统计学的理解，对统计学方法的技巧应用……都有独到之处。

本书第 2 版对第 1 版的材料做了重新安排，并且对许多部分进行了改写，这样读者更易接受和理解。它以更加深入浅出的语言介绍了试验设计的思想和各类设计分析的方法，不仅对工程实际工作者在应用试验设计中有着直接的指导作用，而且为理论研究者提供更加丰富的思想启示。另外，本书第 2 版还引进或强调了本版前言提到的许多东西，还增加了一些新内容，如现代质量控制过程。

本书的翻译工作是集体合作完成的。张润楚、李鹏、刘民千、杨建峰和刘玉坤分别完成了第1~2章、第3~5章、第6~8章、第9~11章和第12~15章翻译初稿，最后由张润楚全面审核、整合、协调定稿，特别是刘民千和杨建峰在最后的修改定稿中发挥了重要作用。

上述关于统计学和试验设计的看法仅是译者的初浅认识。对于本书的翻译，由于我们的水平所限，加上时间紧迫，尽管我们在翻译过程中付出了很大努力，但仍然可能会在译文或其他方面存在不妥甚至错误之处，敬请同行专家和读者给予批评指正。

张润楚

2009年9月于南开大学和东北师范大学

## 第2版前言

在修订这本书的时候，我们深切怀念我们已逝的朋友和同事 Bill Hunter (William G. Hunter, 1937—1986)。我们怀念他的思想、建议和令人愉悦的鼓励。但是，我们相信他的精神已经和我们紧紧地连在了一起，并且他会为我们所做的一切感到高兴。

本书第2版的目的仍然和第1版一样：

1. 使试验者获得实际可行的科学方法和统计工具，这些工具可以大大促进创新、解决问题和获得新发现。

2. 解释这些工具如何使用以及如何与主管专家一起在他们的研究过程中应用。

应该说，使得 Bill 尤为高兴的发展是现今的可喜氛围，这些技术现在在工业中都得到了应用，而且由于目前高速计算机普遍适用性，使得这些广泛的方法只要有必要的话都允许在计算上具备使用的条件。

在如“六西格玛”这样的旗帜下，管理部门已经认识到在经济研究方面培训其工作团队的重要性。现在科学方法的民主化，使许多人已经或正在利用自己的创造力以及还没有来得及实现的才智来寻找对问题的解决和得到新发现。此外，“团队思想”不仅能加速质量的改进，而且通过这种思想容易发现和检验创新的自然领导者并施展其才能让他们去领导。为使得各种各样的创造性工作成为可能，需要把现代哲学和过程改进的方法传授给所有那些处在组织者层面上的人们。我们相信，不管是培训者还是被培训者都会觉得这本书是有用的。而且根据我们长期的经验，相信这本书中的材料对正在大学里需要有这类知识但常常又忽视的那些科学家和工程师来说，也是适合用来做训练的。

本书第2版对第1版的材料做了重新安排，并且进行了改写，使得读者更易于接受和理解。另外，这个新版本还引进或强调了许多新的东西。

对于一个成功的研究者来说，除了统计上的考虑外，还有许多其他方面的问题必须关注。例如下面列举的这些：

- 需要紧密地与主管专家协同工作。
- 注意选择参考集的重要性。
- 作为对更正式方法的一种辅助，重视图技术的价值和作用。
- 为求解多重响应问题注意非活跃（惰性）因子和典则空间的使用。
- 使用“ $\lambda$ 作图”数据变换有时可以大大提高成功的效率。
- 理解误差的传递以及它在设计那些最小地受（或稳健地对）系统成分中的变差影响的产品和过程中的运用。
- 在工业试验中，特别是在受环境影响的稳健产品和过程的设计中，注意裂区（split-plot）设计及类似安排的作用。
- 重视对解决问题的序贯方法特别是试验设计的序贯安排。
- 如何选择最好的跟随试验（follow-up run）。
- 熟悉某些简单装置（例如，一个纸直升机）的手工设计的研究技术。

- 如何把经验变为实际机制.
- 关注随机化和分区组的使用, 以使得分析允许在“好像”是标准假设成立的情形下来进行.
- 注意复杂试验安排的使用, 特别是 Plackett-Burman 设计及其分析.
- 关注设计信息函数.
- 关注导引到过程控制、预测和时间序列分析的知识.
- 注意关于革新过程操作更深入的讨论.

## 关于章末问题和习题的使用

在每章末都有问题和习题, 它们有两个用处: 在阅读这一章之前, 可以把它们作为预习, 以帮助了解关键点和指导阅读, 而在阅读之后, 可以把它们作为实际应用和练习.

当你应用这些思想, 特别是, 如果你遇到异常的成功或者失败, 我们会很感兴趣学习你的经验. 我们尽力要写出一本很有用而且又通俗易懂的书, 但未必完全如意. 如果你有什么关于如何改进这本书的建议, 请写给我们.

## 致谢

我们特别感谢 Ernesto Barrios Zamudio 在本书的准备过程中给予的慷慨帮助. 还要感谢 René Valverde Ventuara、Carla Vivacqua 和 Carmen Paniagua Quiñones 对我们的帮助. 我们还感激 David Bacon Mac Berthouex、Søren Bisgaard、Bob DeBaun、David Steinberg、Dan Meyer、Bill Hill、Merv Muller 和 Daniel Peña, 感谢他们对该书前一版的评论和有价值的建议, 并且感谢给予过我们耐心鼓励的 Claire Box 和 Tady Hunter. 尤其是, 在这长期努力中, 我们幸运地遇到 Lisa Van Horn 编辑, 其与我们愉快的合作和智慧的讨论贯穿整个书的出版准备过程, 在此一并致以谢意.

George E. P. Box  
J. Stuart Hunter

*Madison, Wisconsin*  
*Princeton, New Jersey*  
*April 2005*

# 目 录

译者序

第2版前言

|                             |    |
|-----------------------------|----|
| 第1章 促进知识的产生                 | 1  |
| 1.1 学习过程                    | 1  |
| 1.2 重要的考虑                   | 3  |
| 1.3 试验者的问题及统计方法             | 5  |
| 1.4 一个典型研究                  | 6  |
| 1.5 如何用统计技术                 | 8  |
| 第2章 基础知识(概率、参数和统计)          | 12 |
| 2.1 试验误差                    | 12 |
| 2.2 分布                      | 13 |
| 2.3 统计量和参数                  | 16 |
| 2.4 位置和散布的度量                | 16 |
| 2.5 正态分布                    | 19 |
| 2.6 正态概率作图                  | 22 |
| 2.7 随机性和随机变量                | 23 |
| 2.8 协方差和相关作为线性相依性的度量        | 25 |
| 2.9 学生氏 $t$ 分布              | 26 |
| 2.10 参数估计                   | 28 |
| 2.11 来自正态总体的随机抽样            | 29 |
| 2.12 卡方和 $F$ 分布             | 30 |
| 2.13 二项分布                   | 32 |
| 2.14 泊松分布                   | 35 |
| 附录 2A 观测的线性组合的均值和方差         | 37 |
| 第3章 比较两总体: 参照分布、检验和置信区间     | 42 |
| 3.1 相关的参照集合和参照分布            | 42 |
| 3.2 随机配对比较设计: 男孩鞋子例子        | 51 |
| 3.3 分区组和随机化                 | 57 |
| 3.4 小结: 简单试验中的比较、重复、随机化和分区组 | 58 |
| 3.5 显著性检验再讨论                | 59 |
| 3.6 离散数据的推断: 二项分布           | 66 |
| 3.7 频数(每单位的计数)的推断: 泊松分布     | 69 |

|                                |     |
|--------------------------------|-----|
| 3.8 列联表和关联性检验                  | 70  |
| 附录 3A 两总体比较检验的稳健性对比            | 72  |
| 附录 3B 基于历史数据的参照分布的计算           | 74  |
| 第4章 比较多个总体: 随机区组和拉丁方           | 82  |
| 4.1 在一个完全随机设计中比较 $k$ 个处理       | 82  |
| 4.2 随机区组设计                     | 89  |
| 4.3 裂区试验初步及其与随机区组的关系           | 94  |
| 4.4 多个分区组成分: 拉丁方设计             | 95  |
| 4.5 平衡不完全区组设计                  | 99  |
| 附录 4A 图形方差分析的基础                | 100 |
| 附录 4B 一些有用的拉丁方: 希腊拉丁方和超希腊拉丁方设计 | 101 |
| 第5章 两水平因析设计                    | 105 |
| 5.1 引言                         | 105 |
| 5.2 例1: 三因子(变量)对薄层清晰度的影响       | 105 |
| 5.3 例2: 三因子对某聚合物溶剂的三种物理性质的影响   | 106 |
| 5.4 $2^3$ 因析设计: 试验工厂研究         | 107 |
| 5.5 主效应的计算                     | 109 |
| 5.6 交互作用                       | 110 |
| 5.7 真实的重复试验                    | 111 |
| 5.8 结果的解释                      | 113 |
| 5.9 对照表                        | 114 |
| 5.10 $2^k$ 因析试验中 ANOVA 的误用     | 115 |
| 5.11 观察数据                      | 117 |
| 5.12 多个响应的处理: 宠物食品试验           | 118 |
| 5.13 $2^k$ 因析设计: 工艺过程开发研究      | 122 |
| 5.14 利用正态图和 Lenth 图分析          | 124 |
| 5.15 因析设计数据的其他模型               | 128 |
| 5.16 $2^k$ 因析设计分区组             | 129 |
| 5.17 边做边学                      | 132 |

|                                                                     |     |                                               |     |
|---------------------------------------------------------------------|-----|-----------------------------------------------|-----|
| 5.18 总结 .....                                                       | 135 | 附录 8B 检验方差非齐性的 Bartlett $\chi^2_{\alpha}$ ... | 204 |
| 附录 5A 更大因析设计的分区组 .....                                              | 135 | 第 9 章 变差的多种来源 .....                           | 208 |
| 附录 5B 部分混杂 .....                                                    | 136 | 9.1 裂区设计、方差成分和误差传递 ...                        | 208 |
| 第 6 章 部分因析设计 .....                                                  | 144 | 9.2 裂区设计 .....                                | 208 |
| 6.1 8 次试验中 5 个因子对薄层 6 种特性的效应 .....                                  | 144 | 9.3 估计方差成分 .....                              | 215 |
| 6.2 新产品的稳定性, 8 次试验中 4 个因子, 一个 $2^{4-1}$ 设计 .....                    | 145 | 9.4 误差传递 .....                                | 220 |
| 6.3 一个半拆分的例子: 某种轴承的改进 .....                                         | 146 | 第 10 章 最小二乘与试验设计的必要性 .....                    | 226 |
| 6.4 半拆分的分析 .....                                                    | 147 | 10.1 最小二乘估计 .....                             | 226 |
| 6.5 $2^{7-4}_{III}$ 设计: 一个自行车的例子 .....                              | 149 | 10.2 最小二乘的多样性 .....                           | 234 |
| 6.6 8 次试验设计 .....                                                   | 151 | 10.3 试验设计的起源 .....                            | 246 |
| 6.7 使用表 6.6: 一个例证 .....                                             | 152 | 10.4 非线性模型 .....                              | 253 |
| 6.8 符号交换、折叠反转及序贯安排 .....                                            | 153 | 附录 10A 统计概念的向量表示 .....                        | 255 |
| 6.9 采用多列折叠反转的一项调查 .....                                             | 155 | 附录 10B 最小二乘的矩阵版本 .....                        | 258 |
| 6.10 通过折叠反转将设计的分辨率由 III 提高到 IV .....                                | 158 | 附录 10C 因析设计、搞糟的设计及其他设计的分析 .....               | 259 |
| 6.11 16 次试验的设计 .....                                                | 158 | 附录 10D 未加权最小二乘和加权最小二乘 .....                   | 261 |
| 6.12 $2^5$ 因析设计的节点半复制: 反应堆例子 .....                                  | 159 | 第 11 章 建模、几何及试验设计 .....                       | 272 |
| 6.13 $2^8$ 因析设计的 $2^{8-4}_W$ 节点 $\frac{1}{16}$ 部分 ...               | 162 | 11.1 一些经验模型 .....                             | 274 |
| 6.14 $2^{15-11}_{III}$ 节点设计: $2^{15}$ 因析设计的 $\frac{4}{64}$ 部分 ..... | 163 | 11.2 某些试验设计及设计的信息函数 ...                       | 278 |
| 6.15 构造其他的两水平部分 .....                                               | 165 | 11.3 曲线估计的是否充分 .....                          | 281 |
| 6.16 区组效应的消除 .....                                                  | 166 | 11.4 序贯设计策略 .....                             | 282 |
| 第 7 章 补充的部分因析设计及其分析 .....                                           | 173 | 11.5 典则分析 .....                               | 286 |
| 7.1 Plackett 和 Burman 设计 .....                                      | 173 | 11.6 Box-Behnken 设计 .....                     | 294 |
| 7.2 选择跟随试验 .....                                                    | 181 | 第 12 章 响应曲面方法的某些应用 .....                      | 304 |
| 7.3 对使用部分设计的辩护 .....                                                | 187 | 12.1 重复试验以改进产品设计 .....                        | 304 |
| 附录 7A 技术细节 .....                                                    | 189 | 12.2 通过数据变换简化响应函数 .....                       | 313 |
| 附录 7B PB 设计的一个近似偏分析 .....                                           | 191 | 12.3 探索和研究多重响应数据的活跃因子空间和惰性因子空间 .....          | 316 |
| 附录 7C Hall 的正交设计 .....                                              | 191 | 12.4 探索典则因子空间 .....                           | 318 |
| 第 8 章 因析设计及数据变换 .....                                               | 197 | 12.5 从经验论到机械论 .....                           | 323 |
| 8.1 一个两向(因析)设计 .....                                                | 197 | 12.6 响应曲面法的使用 .....                           | 328 |
| 8.2 来自变换的简化以及增强的灵敏度 ...                                             | 198 | 附录 12A $\hat{\sigma}$ 的平均方差 .....             | 328 |
| 附录 8A 数据变换的基本原理 .....                                               | 204 | 附录 12B 12.3 节中多重响应试验的数据和估计的效应 .....           | 329 |
|                                                                     |     | 第 13 章 设计稳健产品和稳健过程: 引论 .....                  | 336 |
|                                                                     |     | 13.1 环境的稳健性 .....                             | 336 |
|                                                                     |     | 13.2 对组成部分波动的稳健性 .....                        | 343 |

|                                  |     |                                |     |
|----------------------------------|-----|--------------------------------|-----|
| 附录 13A 关于环境稳健性的一个<br>数学阐述 .....  | 347 | 14.6 用模型做预测 .....              | 368 |
| 附录 13B 准则的选取 .....               | 348 | 14.7 干预分析：洛杉矶空气污染的<br>例子 ..... | 371 |
| 第 14 章 过程控制—预测和时间序列：<br>引论 ..... | 353 | 第 15 章 革新过程操作 .....            | 375 |
| 14.1 过程监测 .....                  | 353 | 15.1 多个因子 .....                | 376 |
| 14.2 指数加权移动平均 .....              | 356 | 15.2 多重响应 .....                | 379 |
| 14.3 CuSum 图 .....               | 358 | 15.3 革新过程操作委员会 .....           | 379 |
| 14.4 过程调整 .....                  | 360 | 附录 表 .....                     | 382 |
| 14.5 对一些时间序列模型及应用的<br>简略回顾 ..... | 366 |                                |     |

# 第 1 章 促进知识的产生

## 1.1 学习过程

知识就是力量。关键是如何创新和利用知识。但是知识的获得可能是复杂的，要花费时间和付出代价。为了事业上获得成功，你必须学会如何学习。学习这一概念并不深奥，要紧的是如何使好的思想得以产生、达到一个过程的改进、一个新的稳健产品或工艺得到开发，等等。通过使用这本书，你可以发现如何大大简化和加速知识的产生、检验以及新思想的发展过程。同时你也会发现统计方法(特别是试验设计)是催化新知识产生的科学方法并且可以大大提高效率。

学习是通过如图 1.1 所描述的反复过程而进行的。开始由一个可以用数据来比较的某些必要结果的演绎(deduction)过程引导出来一个初始的思想(模型、假设、理论或者猜想)。当结果和数据不一致时，通过一个叫做归纳(induction)的过程，这种差异可以导致对模型的修正。于是开始在

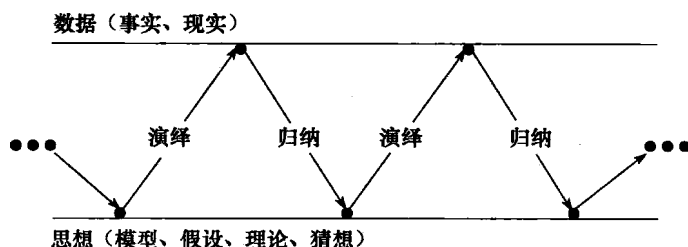


图 1.1 迭代学习过程

迭代中的第二个循环。计算出所修正的模型的结果并且再一次和(旧的或新获得的)数据进行比较，依此下去，可能得到进一步的修改并获得知识。获得过程的数据可以通过科学实验，但也可以是通过到图书馆查询或到互联网上浏览。

### 归纳-演绎学习：日常体验

反复的归纳-演绎过程是人们日常体验的重要部分，它适合于人脑组织结构，并且自古希腊大哲学家亚里士多德(Aristotle)时代就已广为人知了。例如，一个化学工程师 Peter Minerex 每天早晨把他的车停在分配的停车位置。下班后的一天下午，他遵循如下的归纳-演绎学习顺序：

模型：今天和往常每天一样。

演绎：我的车会在它的停放位置。

数据：它不在那个地方。

归纳：一定有人把它开走了。

模型：我的车被人偷了。

演绎：我的车会不在那个停车场地。

数据：不，哦，它就在那边！

归纳：有人把它开走了但又放回来了。

模型：一个小偷开走了但又放回来了。

演绎：我的车可能被损坏了。

数据：它没有被损坏而且没有锁。

归纳：是有人有钥匙开走车的。

模型：我太太用过我的车。

演绎：她可能在车里留了纸条。

数据：是的，你看，纸条就在这儿。

假设要解决一个特定问题并且初始的思索提出某些相关的想法或理论。然后就会寻找数据来进一步支持或者否定这一理论。这可能由以下某些方面组成：你的领域和互联网的搜索、步行到图书馆查询、与合作者和执行者的讨论会，对过程的被动观测或者主动试验。在任何情况下，事实和收集到的数据有时会进一步证实你的猜想，此时你可能已经解决了你的问题。然而，经常发生的是你的初始想法仅有部分是对的或者完全是错的。在这后两种情况下，演绎和现实情况之间的差异使得你还要继续搜集数据。这就可能导致一种修正或者完全不同的想法，并且对你目前的数据进行重新分析或者产生新的数据。

人类的左右脑就是用来执行这种持续的归纳-演绎交流的。当这种反复过程能够解决问题时，你就不会期待这个解的本性或者达到这个解的路径是唯一的。

### 一个化学例子

Rita Stoveing 一个化学家，有过如下的想法：

**模型** 因为一种最新发现的催化剂的某些性质，在一种特定反应混合物中它的存在可能会造成化学药品 A 跟化学药品 B 以高产量形成一种有价值的产品 C。

**演绎** 丽塔有一个试探性的假设并且推断出它的结果，但是没有数据来验证或者否定它的真实性。根据和同事们交谈、小心地查阅文献以及在计算机上的更进一步地搜索，她确认还未曾有人实施操作来完成这一课题。因此，她决定进行一些适当的试验。应用她的化学知识，她要在一个细心选择的条件上做一个试验。特别地，她猜想温度 600℃ 应该值得一试。

**数据** 第一个试验的结果令人失望。所要求的产品 C 是一种无色、无味的液体，而目前所得到的是一种黑色焦油状的产品，且只含不到 1% 的物质 C。

**归纳** 在这点上原始模型和数据不一致。这个问题让丽塔感到烦恼并且晚上在她的丈夫 Peter Minerex 面前表现得有些不高兴。但是在第二天早晨洗澡的时候，她开始沿下面的思路想：产品 C 可能首先已经以高产量形成，但然后它又被分解。

**模型** 理论提出的反应条件太严格。

**演绎** 稍低些的温度可能得到 C 的满意产量。首先将反应温度降至 550℃ 然后又降至 500℃ 来做两个进一步的试验。

**数据** 从这两个试验得到的产品减少了焦油状而且也不那么黑了。在温度 550℃ 的试验中得到产品 C 的 4%，而在温度 500℃ 的试验中得到 17%。

**归纳** 给定这些数据并根据她对该反应的理论知识，她决定应该进一步做试验，不只是考虑温度，而且考虑一些其他因子（例如，浓度、反应时间、催化剂量）并研究产品的其他特征（例如，各种纯度、黏性水平）。

为了经济地评价这样的复杂系统，她将要设计设计的试验和统计分析。在本书的后面你会看

到如何利用统计工具进行后续的研究。

**练习 1.1** 根据自己的研究领域，如工程、农业、生物、基因学、教育、医学、心理学等，描述一个真实的或者设想的反复学习的例子。

### 一个反馈回路

在图 1.2 中的归纳—演绎迭代表现为一个反馈过程。一个初始的思想(假设或模型)用  $M_1$  表示在图的左边。通过演绎，考虑  $M_1$  的期望结果——如果  $M_1$  成立可能发生什么而  $M_1$  不成立可能发生什么。还要推演一下需要什么样的数据来探究  $M_1$ 。你打算要做的试验计划(设计)在这里用一个框架来表示，通过这个框架可以看到特性的真实状态的某些方面。注意，当你进行一次试验时这个框架是在你的选择下的<sup>⊖</sup>(即你能控制这个框架)。产生的数据表现(或多或少被“噪声”(noise)即试验误差弄模糊的)特性真实状态的某个方面(虽然不总是相关的)。分析的数据可以与  $M_1$  的期望(推演)结果进行比较。如果一致，你的问题可能得到解决。如果不一致，那么它们不一致的情况可能启示你去发现你的原始想法  $M_1$  的问题及需要如何去修改。利用同样的数据，可以考虑另外的分析以及原始模型可能的修改  $M'_1$ ,  $M''_1$ , ...。可以弄清楚你的原始思想是否是错的或者至少要做相当地修改。这可能要求选择一个新的或者扩充的试验设计去阐明该特性的状态的可能不同的方面。这可能会使该问题得到一个满意的解决，或者提供一些线索提示下一步最好如何去进行。

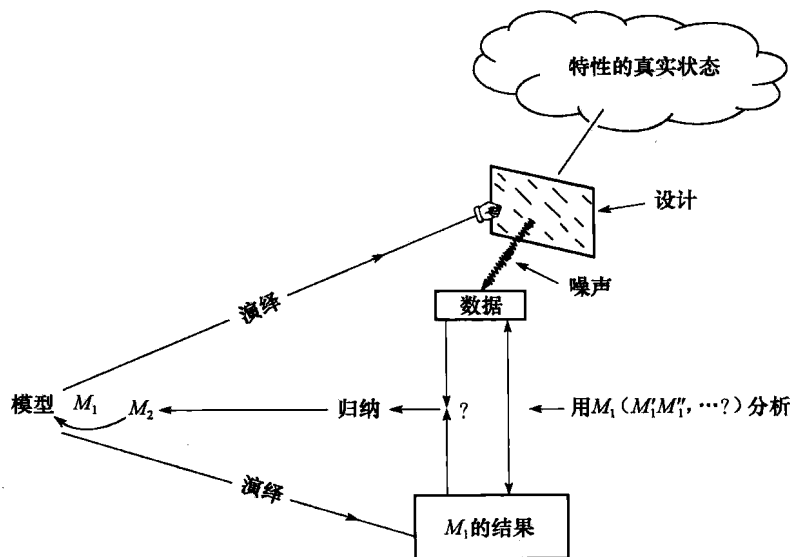


图 1.2 解一个视为反馈回路的迭代问题

## 1.2 重要的考虑

### 主题知识

注意，主题知识对感知和探索尝试性模型以及知道到哪里去寻求帮助的重要性。

⊖ 当然这是不真实的，对于偶然事件你是不能控制的。

### 对问题解答的路径不是唯一的

当 Peter Minerex 一开始注意到他的汽车不见的时候, 他可能很容易作出不同的行为反应。例如, 他可以立即给警察打电话, 从而启动不同的发现路径(但可能不同等地有效)。类似地, 在化学研究中, 不同的研究者在得到失望的结果之后, 可能决定去探索完全不同的化学途径来得到所期望的产品。其目标是会聚(converge)到一个满意的解——出发点和路径(有时还包括解的特性)对于不同研究者来说会是不同的。

用一个 20 个问题的游戏来解释这些。在这个游戏中, 目的是用不超过 20 个问题来确定一个未知的对象, 其中每个问题都只能从两个不同答案中选择一个。假设要猜的对象是 Abraham Lincoln(亚伯拉罕·林肯)的炉管帽子(stove pipe hat)。开始的思路是与动植物有联系的物品。对瞄准这个思路的游戏竞赛队, 该游戏可能如下进行:

| A 队          |     |                  |     |
|--------------|-----|------------------|-----|
| 问 题          | 答 案 | 问 题              | 答 案 |
| 是联系到动物和人类吗   | 是的  | 是不是本世纪的          | 不是  |
| 是男的还是女的      | 男的  | 19 世纪还是 18 世纪    | 19  |
| 是著名的人还是不著名的人 | 著名的 | 是与内战有联系吗         | 是的  |
| 是搞文学的吗       | 不是  | 是亚伯拉罕·林肯吗        | 是的  |
| 是政治家吗        | 是的  | 这件物品是亚伯拉罕·林肯的帽子吗 | 是的  |
| 美国还是其他国家     | 美国  |                  |     |

但对另一个游戏竞赛队, 这个游戏沿着几乎确定的不同路径进行。例如:

| B 队             |     |              |     |
|-----------------|-----|--------------|-----|
| 问 题             | 答 案 | 问 题          | 答 案 |
| 这件物品是用的吗        | 是的  | 是戴在头上还是别的地方  | 头上  |
| 它是一件服装用品吗       | 是的  | 是一顶著名的帽子吗    | 是的  |
| 是男人用的还是女人用的     | 男的  | 是温斯顿·邱吉尔的帽子吗 | 不是  |
| 是戴在上面的还是系在下面的带子 | 上面的 | 是亚伯拉罕·林肯的帽子吗 | 是的  |

这个游戏是遵从图 1.1 和图 1.2 的迭代模式进行的, 其中的“设计”就是问题的选择。在每一步的猜想, 逐渐地加细, 得到一个能引出新数据的合适问题的选择, 它依次引导至猜想的合适修正。A 队和 B 队沿着不同的路径, 但每个都引导出了正确的答案, 因为数据都是由事实产生的。

玩这个游戏所需要的要素是(a)主题知识和(b)策略知识。关于策略, 众所周知在每一步询问的问题应该是使其划分非以前排除的客体成等概率的两半。在这些例子中, 游戏者通常是试图以像“是男的还是女的?”或“是戴在上面的还是系在下面的带子?”这样的问题来做到这点的<sup>①</sup>。

① 从包含百万个单词的一本字典中任选一个词可以用玩 20 个问题的游戏找出来。首先问“它是在这字典的前一半还是后一半?”如果答案是“前一半”, 然后问“它是在这一半中的前一半还是后一半?”等。注意  $2^{20} \gg 10^6$ 。

策略知识平行于在科学研究中的统计方法的知识。注意，没有策略知识你总能玩这个游戏，虽然多半玩得不好，而没有主题知识就一点也不能玩了。然而，到目前为止，最好是应用主题知识和策略知识这两者。没有统计有可能进行一项研究，但如果没有主题知识就不可能做研究。但是，通过用统计方法会加速趋向于问题的解并且一个好的研究者会变成一个更好的研究者。

### 1.3 试验者的问题及统计方法

摆在研究者面前的三个问题是复杂性(complexity)、试验误差(experimental error)和相关与因果关系(correlation and causation)之间的区别。

#### 复杂性

在对于过程<sup>①</sup>的改进和发现新东西的试验中，通常有必要同时考虑许多“输入变量”，像温度、进料比率、浓度和一组输出变量(如产量、纯度和花费等)上的催化剂等。我们称可控制的输入变量为因子(factor)，称输出变量为响应(response)。在研究关于如何改进一个过程的问题中主要的问题是：“从什么到什么是为了做什么？”具有 $k$ 个因子和 $p$ 个响应时，需要考虑 $k \times p$ 个对象。进而，当因子的一个确定子集(例如温度和压力)可能对改变一个响应(例如产量)有用时，一个完全不同的或者很像是重叠的子集(例如温度和浓度)可能影响着一个不同的响应(例如纯度)。为了得到一个满意的高产量和足够的纯度，折中的方法可能是必要的。还有，某些因子在一个特定响应上可能是以交互作用的方式发挥它们的影响的。例如，由温度的一个特定变化引起产量的变化，而这种变化可能在不同的浓度上又是不一样的。对试验者来说，要把这些事情同时地考虑进去无疑是一个挑战性的问题。挑选、尝试猜测以及使用“一次改变一个因子”的方法从试验原理上看都不大可能很快并经济地得到好的结果。

统计试验设计的使用使之成为可能，它能最小化试验误差的影响，在试验里同时考虑多个因子，并且获得它们是如何单独和共同地对响应行为表现的一个清楚画面。这样的理解可以得出问题的经验解，但是它还能做得更多。拥有来自进行得好的试验结果的主题专家可以沿着如下的线路来推理：“当我看见 $x_3$ 对 $y_1$ 和 $y_2$ 的影响以及 $x_1$ 和 $x_2$ 交互地影响 $y_3$ ，这就提示接着将发生什么，并且我想我们现在应该做的是这个。”理论上的理解可以从经验上的陈述得到启发和飞跃。

#### 试验误差

不能由已知的影响所解释的可变性称为试验误差。因为某些试验误差是不可避免的，知道如何去应付它是本质的。经常地，只有一少部分试验误差可以归于测量方面的误差。在原材料、抽样和试验因子的设置安排方面的差异常常占有更大的成分。好的试验设计会帮助保护真实的效应而不受试验误差的干扰，相反地，有研究者错误地相信那些并不存在的效应。

明智地使用统计试验设计可以大大降低试验误差的混淆影响<sup>②</sup>。另外，统计分析可提供研究中所要估计量的精度的测度(例如在均值或变化比率中的差异)。这就使得有可能来判断对真实效应的

① 过程一词在这里是在一般意义下用的。于是一个过程可以是一种分析方法或者一个产品的制造。

② 这句话的另一种表述是设计的试验可以大大地提高信号—噪声比。

存在性是否存在可靠的根据, 并因此大大增加研究者沿着一个真实而不是错误的路径走的概率。

### 相关性和因果关系的混淆

图 1.3 显示的是奥尔登堡市民人口总数与 7 年每个年末观测到的鸛的数目的关系<sup>①</sup>。虽然在这个例子里几乎不会导致假设: 鸛的数目的增加引起观测到的人口增加, 研究者有时在其他场合为这类错误而愧疚。两个变量  $Y$  和  $X$  之间的相关性是经常发生的, 因为它们两者都与第三个因子  $W$  相联系。在这个鸛的例子里, 因为人口总数  $Y$  和鸛的数目  $X$  两者在 7 年期间都是增加的, 在这种情形共同的因子  $W$  是时间。

**练习 1.2** 给出其他的例子, 其中相关关系存在但因果关系不存在。

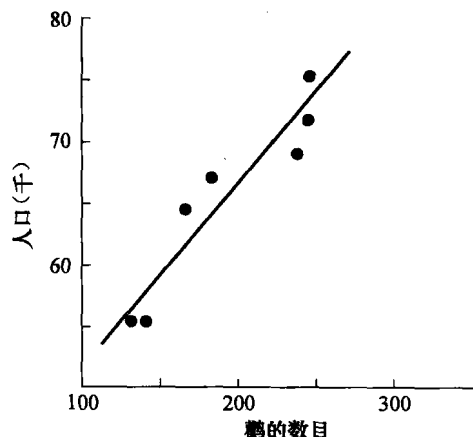


图 1.3 鸛的数目与奥尔登堡 (Oldenburg) 人口关系

## 1.4 一个典型研究

为解释迭代学习过程并同时提供对本书所讨论的问题的一个预览, 我们用一个假想的目的在于提高饮水质量的研究。我们的研究者是化学家 Pita Stoveing 和化学工程师 Peter Minerex。正如你随后要读到的, 考虑同样的统计问题是如何使研究者在任何其他试验科学中都可能面临的。

### 问题

在这个行星上可用的水显然仅有一个有限的量并且我们必定是利用和再利用。下面的研究是必要的, 因为一种特殊的水补给包含一种不能接受的高水平硝酸盐。Minerex 和 Stoveing 已经研制了一种吸附硝酸盐的离子交换树脂。这种新的树脂特别引人瞩目的特点是它对硝酸盐有特别效用并且用这种树脂比目前可用的树脂便宜很多并且可再生。不幸的是, 它只能在实验室条件下在试验量之内做出来。已经认识到在一个商业可行产品能进行生产之前, 一批量的进一步试验是必要的。就他们的研究进程而言, 下面的一个梗概反映他们将如何被引导去考虑不同的问题和不同程度地利用各种不同的统计技术<sup>②</sup>。

### 研究的迭代循环

丽塔和皮特知道在明显类似的条件下, 记录下来的观测可能有相当大的变化。于是, 在开始研究之前, 他们就要努力稳定程序和减少过程变差。而且, 他们期望有大量的不同因子加入到试验并看到各种不同的响应。为使这一研究有效和经济, 他们将需要广泛地应用统计方法。

① 这些数据覆盖 1930—1936 这些年。见 *Ornithologische Monatsberichte*, 44, No. 2, Jahrgang, 1936, Berlin, and 48, No. 1, Jahrgang, 1940, Berlin, and *Statistics, Jahrbuch Deutscher Gemeinden*, 27—22, Jahrgang, 1932—1938, Gustas Fischer, Jena。我们感谢 Lars Pallesen 提供了这些参考。

② 这个假想的研究有一个好处是, 在相继的迭代过程中, 它用到了本书中讨论的大部分技术, 而且在顺序上近似地接近书中出现的顺序。当然, 这仅仅是一个教学法上的设计。然而, 许多研究都要有统计方法贯穿于不同阶段, 有点像在这里所描述的方式。

实况说明 两个研究者每人用了—个统计过程但目前都有些不满意。

### 迭代 I

问题 从哪里可以找到一个快捷、梗概的初等统计原则?

设计和分析 第 2 章: 基础

发现 学习这一章提供的必要知识和评述并为随后的工作做准备。

### 迭代 II

实况说明 Minerex 相信, 而 Stoveing 不相信, 用—种纯度很高(或者说更昂贵)的树脂会改进试验的效果。

问题 如何来比较他们“普通的”树脂和更昂贵的高纯度树脂的效果呢?

设计和分析 第 3 章: 比较两个总体

发现 他们的更昂贵的高纯度树脂大致和普通的树脂效果—样。(Stoveing 是对的!)

### 迭代 III

实况说明 Minerex 对高纯度树脂的想法错了, 但是他们仍然认为普通的树脂会有希望。他们现在决定把他们实验室新的树脂样本和商业上实用的 5 种树脂比较—下。

问题 如何来把他们这种新的普通树脂和商业上实用的 5 种树脂进行比较呢?

设计和分析 第 4 章: 比较多个总体

发现 他们实验室这种新树脂样本和商业上实用的任何一种树脂—样地好, 而且好像稍微好—点。

### 迭代 IV

实况说明 已经表明新的树脂可能做得和它的竞争者—样好。然而, 在预期的经济制造条件下除去硝酸盐方面还不足以达到饮用水要求的标准。

问题 什么是影响消除硝酸盐的最重要的因子? 在影响像流动速率、河床深度和再生时间这些因子的目前制造设备条件下, 能够通过修改来达到对消除硝酸盐的改进吗?

设计和分析 第 5 章、第 6 章、第 7 章和第 8 章: 学习用因子和部分因析设计

发现 用适当的设备修改, 可以达到足够低的硝酸盐水平。

### 迭代 V

实况说明 公司现在断定这种新树脂的生产是可能的而且是有利可图的。为了解更多, 建立了—个引水车间。

问题 如何设置影响这种新树脂的质量和费用的过程因子? 什么是最好的设置?

设计和分析 第 10 章、第 11 章和第 12 章: 最小二乘、多维建模和响应曲面方法

发现 引水车间研究表明, 用适当的过程因子的设置在一个合理的费用下可以生产出来质量令人满意的产品。

### 迭代 VI

实况说明 在推荐这个生产过程之前, 抽样和检验的问题还得要解决。

问题 抽样和检验的方法如何进行细化以给出对新树脂特性的可靠判断?

设计和分析 第 9 章: 变差的多种来源

发现 在产品的抽样和化学分析中, 识别和测量出变异性成分。利用这—信息, 设计出—个抽样和检验协议草案, 以最小的花费最小化判断的差异。