

外国优秀统计学教材系列丛书 · 翻译版



作者

阿格雷斯特 (Alan Agresti)

翻译

陈家鼎

陈奇志

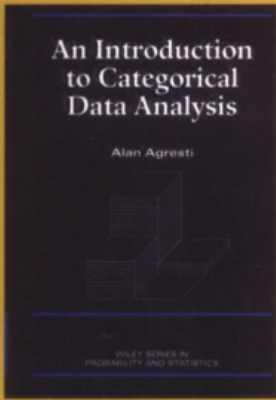
审校

孙山泽

全国统计教材编审委员会组织引进

类别数据分析导引

An Introduction To Categorical Data Analysis



<http://www.wiley.cn>



定价: 36.00 元



0212. 1/36

2008

外国优秀统计学教材系列丛书· 翻译版

全国统计教材编审委员会组织引进

类别数据分析导引

————— An Introduction to Categorical Data Analysis

作者 Alan Agresti
翻译 陈家鼎
陈奇志
审校 孙山泽

 中国统计出版社
China Statistics Press

Library of Congress Cataloging-in-Publication Data:
Agresti, Alan.
An introduction to categorical data analysis/Alan Agresti

ISBN 0-471-11338-7

Copyright © 1996 by John Wiley&Sons, Inc.
All rights reserved. Published simultaneously in Canada

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

本书中文简体字版由 John Wiley&sons, Inc 授权中国统计出版社独家出版。未经出版者书面许可, 不得以任何方式翻印、拷贝、仿制或转载。

版权所有, 侵权必究。

北京市版权局著作权合同登记号: 图字: 01-2006-0656

(京)新登字041号

图书在版编目(CIP)数据

类别数据分析导引: 翻译版 / (美)阿格雷斯特(Agresti, A)著; 陈家鼎, 陈奇志译.

北京: 中国统计出版社, 2006.6

(外国优秀统计学教材系列丛书. 翻译版)

ISBN 987-7-5037-5021-2

I. 类... II. ①阿...②陈...③陈... III. 统计数据-统计分析(数学)-教材 IV. 0212.1

中国版本图书馆CIP数据核字(2006)第049069号

责任编辑 / 吕军

装帧设计 / 艺编广告

出版发行 / 中国统计出版社

通信地址 / 北京市西城区月坛南街57号

邮政编码 / 100826

办公地址 / 北京市丰台区西三环南路甲6号

电 话 / 邮购(010)63376907 书店(010)68783172

印 刷 / 北京顺义兴华印刷

经 销 / 新华书店

开 本 / 850 × 1092 毫米 1/16

字 数 / 300 千字

印 张 / 18

印 数 / 1—3000 册

版 别 / 2006年6月第1版

版 次 / 2008年3月北京第1次印刷

书 号 / ISBN 987-7-5037-5021-2/F·2344

定 价 / 36.00 元

中国统计版图书, 如有印装错误, 本社发行部负责调换。

出版说明

21 世纪的竞争是人才的竞争，是全球性人才培养机制的较量。

如何培养面向现代化、面向世界、面向未来的高素质的人才成为我国人才培养的当务之急。为此，教育部发出通知，倡导在全国普通高等学校中使用原版外国教材，进行双语教学，培养适应经济全球化的人才。

为了响应教育部的号召，促进统计教材的改革，培养既懂统计专业知识又具备较高英语语言能力的统计人才，全国统计教材编审委员会在国家统计局领导的大力支持下，组织引进了这套“外国优秀统计学教材”。

为了做好“外国优秀统计学教材”引进工作，全国统计教材编审委员会将其列入了“十五”规划，并成立了由海内外统计学家组成的专家委员会。在对国外统计学教材的使用情况进行了充分了解，对国内高等院校使用外国统计学教材的需求情况进行了仔细分析，并对从各种渠道推荐来的统计教材进行了认真审定的基础上，制定了引进教材书目。在确定引进教材书目的过程中，我们得到了国内外有关专家、有关院校和外国出版公司及其北京办事处的支持和帮助，在此致谢。中国人民大学统计学系的吴喜之教授不仅推荐了大量的优秀候选书目，而且校译了影印教材的翻译目录，为这套教材的及早出版作了大量的工作，我们表示衷心的感谢。

这套引进教材多数是国外再版多次、反响良好，又比较适合国内情况、易于教学的统计教材。我们希望这套引进教材的出版对促进我国统计教材的改革和高校统计学专业双语教学的发展能够起到重要的推动作用。

全国统计教材编审委员会

2002年8月28日

译者序

类别数据(Categorical data)也称分类数据、属性数据、定性数据,乃是关于类别的计数数据,这里“类别”是指要考察的事物可划分为有限个不同的类,每次观测时事物出现在一个类里。例如,性别(男,女),人种(黄种人,白种人,黑种人,其他人种),痛感的程度(不痛、轻微、严重),对一项提案的态度(赞成、反对、尚未确定)等都是类别。即使是连续变量(其值可以是某个区间中的任何数)有时也可以(或有必要)将其值划分为有限个不同的类别。“类别”在各种实际问题中出现,因而类别数据分析的方法有广泛的应用,特别是在社会科学和医学领域。A. Agresti的这本书系统地介绍了类别数据分析的概念和方法,有下列特点。

1)突出统计方法。针对各种类型的实际问题介绍有关的概念和方法,基本上不进行数学推导。从直观的角度提出应怎样进行分析及要注意的事项,而不去论述这些方法背后的数学定理的证明。

2)实用性强。全书有上百个含有具体数据的实际问题的例子,这些例子来自很多领域(主要是社会科学和医学),对读者学会理论联系实际有很大的启发、借鉴作用。

3)要求的数学知识少。不要求读者学过微积分和矩阵代数,只要求读者具有初等的概率统计知识。

4)强调使用统计软件。全书不提供复杂的计算公式和计算步骤,而是要求读者学会使用SAS或SPSS软件进行计算。在附录里结合各章出现的方法和数据介绍这些软件的使用方法。这符合当今学用统计方法的时代潮流。

这本书也存在不足之处。主要是行文比较啰唆。为了避免数学上的抽象,作者往往不用公式而用一大堆话语说明一些公式要表达的内容,夹叙夹议,但是有些内容不用公式很难使读者得到准确的理解。读者在阅读各章时,应注意从具体问题到科学的抽象再回到具体问题的解决,以这种思维方式对内容进行概括。

为了进一步掌握类别数据分析的方法和有关的统计理论,读者可去看下列著作:

Agresti, A. (1990). *Categorical Data Analysis*, John Wiley.

Santner, T. J. and Duffy, D. E. (1989). *The Statistical Analysis of Discrete Data*, Springer-Verlag.

Bishop, Y. M. M et al. (1975). *Discrete Multivariate Analysis: Theory and Practice*. (中译本《离散多元分析:理论与实践》,张尧庭译,中国统计出版社,1998)。

这三本书内容丰富,但要求读者具有高等数学和概率统计的基础理论知识并受过严格的数学训练。

本书译稿承蒙孙山泽教授进行审校。他提出了宝贵的修改意见并帮助译者进行修改。我们在此向他表示感谢。

限于译者的水平,译文肯定还有不妥之处。欢迎专家和读者批评指出。

陈家鼎 陈奇志
北京大学 数学科学学院
光华管理学院

序 言

近些年来,类别数据(Categorical data)的专门统计分析方法的使用(特别是在生物医学和社会科学上的应用)迅猛增加。这部分地反映出,分析类别数据的精致方法在过去几十年里有了发展。也反映出,科学家和应用统计学家在科学方法上的素养不断提高,他们中的大多数人已经很清楚:使用针对连续数据的方法去处理类别数据不仅是不必要的而且常常是不合适的。

本书提供分析类别数据的最重要方法。这些方法在长时期里起着显著的作用,例如卡方(χ^2)检验和相关性度量。但是叙述的重点是罗吉斯提克(Logistic)回归和对数线性模型技巧。这些方法与连续变量的线性模型方法有很多相同的面貌。

本书的要求较低,不要求读者熟悉微积分和矩阵代数等高等数学。读者应该具有这样的背景:懂得为大学生或非统计专业的研究生开设的两个学期的统计方法课的内容。这些背景知识应包括:估计,假设检验,回归建模和方差分析。

本书是为学习类别数据分析入门课程的大学生编写的,但是也是为涉及类别数据分析的应用统计学家和应用科学家编写的。我希望本书能帮助人们处理来自社会科学、行为科学和医学科学的类别响应数据以及来自公共卫生学、市场学、教育、生物学与农业科学和工业质量控制领域的类别响应数据。

第一章至第六章的材料构成类别数据分析入门课程的核心部分。第二章和第三章叙述了列联表的标准描述和推断方法,例如优势比,独立性检验和 Cochran-Mantel-Haenszel 方法。但是,我觉得只有通过了解统计方法的背景才能加深对方法的理解。这样,叙述的重点放在类别响应的建模上。第四章对二项数据和计数数据介绍了广义线性模型。第五章就二项数据讨论了罗吉斯提克回归。第六章就泊松数据(计数数据)讨论了对数线性模型。后面各章把这些模型联系起来并扩展到多种问题上去。我相信,罗吉斯提克模型比对数线性模型更为重要,因为在类别响应的大多数应用中类别变量是单个二值或多值响应变量。因此,我对罗吉斯提克模型给予更多的注意,这与大多数入门书有所不同。

我喜欢在讲授类别数据时,把罗吉斯提克回归、对数线性模型同通常的回归及方差分析(ANOVA)统一起来。第四章就是在广义线性模型的大伞下这样做的。某些教师愿意轻快地经过这一章,主要是用来针对二项数据介绍罗吉斯提克

回归与针对泊松数据介绍对数线性模型。教师们可以选择第七章至第九章的某些节作为头六章中某些题材的补充。第七章包括了与罗吉特(Logit)模型和对数线性模型的构造有关的题材,如图表示,有序信息的利用,稀疏数据的影响。第八章介绍多项响应(包括名义的和有序的)变量的罗吉特模型。第九章叙述成对数据的分析方法。主要关心生物统计方向的读者可把较大力量放在第八章和第九章上而不是第六章和第七章上,而主要关心社会科学方向的读者则反过来,较大精力放在第六章和第七章上。在本书中,附有*号的标题表示相应的节和小节在首次介绍时是不重要的。

这本书在技术难度上低于我的书《类别数据分析》(Categorical Data Analysis, Wiley 1990)。我希望本书对于那些喜欢有更多应用性介绍的读者是有吸引力的。例如,本书不试图推导似然方程,不证明渐近分布,不讨论当前的研究工作,也不提供完全的文献目录。这本新书的某些特点是1990年的那本书所没有的。例如,在附录里介绍了如何使用用于广义线性模型的新的SAS方法(几乎包括了本书中所提出的各种方法的操作步骤),还有关于相关图的一节(7.1节)以及介绍各方法发展的历史整体观点的一章(第十章)。此外,本书保留了1990年那本书的很多特点,包括

1. 基于广义线性模型的统一表述
2. 有序数据的方法
3. 精确的小样本方法
4. 多类别响应模型
5. 成对比较的方法
6. 有200多个练习题
7. 约有100个具有实际数据的例子

本书中提供的大多数方法都需要大量的计算。对于大部分内容,我避免了复杂计算的细节,我觉得计算机软件能够代替这种单调乏味的计算工作。在很多大的商业软件包里,都有类别数据分析软件可供使用。我建议本书的读者使用软件来回答家庭作业问题和核查书中的例子。附录里就类别数据分析讨论了SPSS并对本书中讨论过的几乎所有方法叙述了使用SAS(特别是PROC GENMOD)的例子。为了参考方便,这些方法是按其在本书各章出现的形式编排的。附录中的表和书中分析过的很多数据集均是从Stat Lib得到的。这个文件可在World Wid Web at <http://Lib,Stat.Cmu.edu/datasets/agresti>. 上利用。

我感谢对本书的手稿进行过评论或对本书的例子和材料提出过建议的许多人。这些人有: Patrica Altham, Scott Beck, James Booth, Jane Brockmann, Brent Coull, Al DeMaris, Harry Khamis, Svend Kreiner, Stephen Stigler, Joffre Swait, Larry Winner, 以及一位匿名的书评作者。

我也要特别感谢 John Hilton 和 Peter Imrey, 他们花了很多时间对初稿进行了

评论并提出了多项有帮助的意见。Atalanta Ghosh 帮助我得到某些数据集, Brent Coull 用图形软件帮助我。还要感谢 Wiley 公司里一直关心本书的 Roach 和给予高规格支持的其他职员。当然,我也要感谢我的妻子 Jacki Levine, 她提出了许多意见并经常鼓励我。

最后要说的是,本人的职业生涯带来了结识世界上许多人的机会,很多专业同行变成了很好的朋友。谨以此书奉献给他们。

Alan Agresti

英国伦敦

1995 年 10 月

目 录

序言

第一章 引言 1

- 1.1 类别响应数据 1
- 1.2 抽样模型 3
- 1.3 对比率的推断 7
- 习题 11

第二章 二维列联表 15

- 2.1 列联表的概率结构 15
- 2.2 比较 2×2 表中的比例 18
- 2.3 优势比(优比) 20
- 2.4 独立性的 χ^2 检验 25
- 2.5 有序数据的独立性检验 30
- 2.6 小样本情形的精确推断 35
- 习题 40

第三章 三维列联表 47

- 3.1 部分关联性 47
- 3.2 Cochran-Mantel-Haenszel 方法 53
- 3.3 关于条件关联性的精确推断* 57
- 习题 59

第四章 广义线性模型 63

- 4.1 广义线性模型的成分 63
- 4.2 二值数据的广义线性模型 65
- 4.3 计数数据的广义线性模型:泊松回归 71

4.4	模型的推断与模型的检查	77
4.5	拟合广义线性模型*	82
	习题	85
第五章	罗吉斯提克回归	91
5.1	罗吉斯提克回归模型的意义	91
5.2	罗吉斯提克回归的推断	96
5.3	模型的检查	98
5.4	与定性的预测因子相应的罗吉特(logit)模型	104
5.5	多重罗吉斯提克回归	108
5.6	罗吉斯提克回归的样本量和功效*	114
5.7	罗吉斯提克回归的精确推断*	116
	习题	119
第六章	列联表的对数线性模型	129
6.1	二维表的对数线性模型	129
6.2	三维表的对数线性模型	133
6.3	对数线性模型的统计推断	137
6.4	高维的对数线性模型	140
6.5	对数线性模型与罗吉特模型的关系	144
	习题	148
第七章	罗吉特模型和对数线性模型的建立与应用	154
7.1	关联图与可折叠性	154
7.2	对有序关联性的建模	159
7.3	条件独立性的检验	164
7.4	稀疏数据的影响	168
7.5	模型拟合的一些细节*	172
	习题	176
第八章	多类别的罗吉特模型	182
8.1	名义响应变量的罗吉特模型	182
8.2	有序响应变量的累积罗吉特模型	188
8.3	有序响应的成对类别罗吉特*	191
	习题	195
第九章	成对比较的模型	201
9.1	相依比例的比较	202

9.2	成对数据的罗吉斯提克回归	204
9.3	方表的对称性模型和拟对称性模型	208
9.4	比较边缘分布	212
9.5	分析评定者的一致性*	215
9.6	成对选择的 Bradley-Terry(布拉得雷-特立)模型*	220
	习题	222
第十章	20 世纪类别数据分析的历史概观*	230
10.1	Pearson 与 Yule 的关联性争论	230
10.2	Fisher 的贡献	232
10.3	罗吉斯提克回归和对数线性模型	233
10.4	晚近的发展	236
附录	用 SAS 和 SPSS 进行类别数据的分析	238
	χ^2 分布的上侧分位数表	250
	参考文献目录	251
	例子索引	254
	名词索引	259

1

引言

从评定新的医学处理(例如医疗措施)的价值到评估在各种有争议的问题上影响我们看法的因素,今天的科学家正在发现分析类别数据的方法有无数的用处。这本书正是为了这些科学家及其合作共事的统计学家和正准备从事此类工作的其他人员而写作的。本书提供了分析类别数据的方法的应用性导引,在叙述上强调方法的使用和解释而不是方法背后的理论。

第一章介绍基本的统计预备知识。温习了类别类据的两个关键性的概率分布——二项分布和泊松(Poisson)分布。还介绍了最大似然法(对类别数据估计参数的最流行方法)及其在分析比例数据时的用处。我们先来讨论类别数据的主要类型并阐明本书的纲要。

1.1 类别响应数据

首先给出类别数据(Categorical data)的定义。类别变量是这样的变量,其值是类别,全部值乃是一些类别组成的集合。例如“政治态度”可分为三个类别:自由主义的,温和主义的,保守主义的。早餐食品的选择可划分为三个类别:热的,冷的,其他。阿尔茨海默(Alzheimer)疾病的诊断测试可使用类别:有症状,无症状。有且只有一个类别适用于一个个体,而不是类别的集合。

类别划分(或称类别尺度)已渗透进社会科学里用来刻画人们对各种问题(或事务)的态度和意见。类别尺度也常常出现在医学里用来刻画种种情况,例如病人手术后是否活着(是,否),受伤的程度(未受伤,轻度,中度,重度),疾病的阶段(初期,后期)。

虽然类别变量在社会科学和医学里很普遍,但类别变量决不限于这两个领域。类别变量在下列领域中也很有见。行为科学(例如精神病人的诊断类型:精神分裂症,抑郁症,神经官能症),公共卫生学(例如是否由于对艾滋病了解已引起安全套的使用增加(是,否)),动物学(例如,考察短吻鳄的主要食物爱好的三个类别:鱼,无脊椎动物,爬行类动物),教育(例如学生对一项考试题的回答分为两类:

正确,不正确),市场营销(考虑消费者的偏好,一种商品有三种主要款式:A,B,C)。甚至在高度定量化的领域(如工程科学和工业质量控制),当产品(或项目)按是否符合某种标准来分类时,也要出现类别变量。

1.1.1 响应变量和解释变量的区分

很多统计分析都区分响应变量和解释变量。例如,回归模型描述连续变量(例如年收入)的分布如何受解释变量(例如受教育的年限和工作经历的年限)的影响而变化。响应变量有时叫因变量或变量 Y ,而解释变量有时叫自变量或变量 X 。

本书的主题是,讨论响应变量为类别变量时如何进行分析。在前一小节里罗列的类别变量(例如政治态度)都是响应变量;在某些研究工作中这些类别变量也可成为解释变量。类别响应变量的统计模型用来分析这样的响应是如何受解释变量影响的。例如,可以研究人的政治态度如何受年收入、受教育的程度、与宗教的关联、年龄、性别、种族等因素的影响。解释变量可以是类别变量,也可以是连续变量。

1.1.2 名义与有序尺度的区别

类别变量有两种刻画尺度。很多类别的划分有自然的次序。例如,对堕胎的合法性所持的态度可划分为三类:一概不批准,在某些情况下批准,一概批准。对公司的储备水平的评价可分为三类:太低,基本合适,太高。对医学处理(治疗)的反应可划分为四类:极好,好,尚可,差。对一个人是否精神上有病的诊断可划分为四类:肯定有,很可能有,不大可能有,肯定没有。有顺序尺度的类别变量叫做有序变量。

没有顺序尺度的类别变量叫做名义变量。例如,与宗教的关系(天主教徒,犹太教徒,新教徒,等等),上班的交通方式(汽车,自行车,公共汽车,地铁,步行),最喜欢的音乐类别(古典音乐,乡村音乐,民族音乐,爵士音乐,摇滚音乐)。住宅的类别选择(公寓,一套公寓房间,单独的房屋,其他)。

对于名义变量,登录类别的顺序是无所谓的,而且统计分析不依赖于这种顺序。对名义变量制定的分析方法,不管类别的登录顺序如何,应给出同样的结果。对有序变量制定的分析方法是依赖于类别的顺序的。我们登录类别是从低到高或从高到低通常是无所谓的。但是,若用其他方法对类别另行排序,则有序分析的结果将可能不同。

对有序变量制定的分析方法不能用到名义变量上去,因为名义变量没有有序类别。对于名义变量制定的分析方法可以用于名义变量或有序变量,因为这样的方法只是对类别进行分析,不管这些类别是否有序。但是这样做对于有序变量来说没有利用“序”带来的信息,将造成效率上的损失。通常对于类别的实际划分(具体尺度)最好采用合适的分析方法。

类别变量常称为定性变量,以便与数值变量或定量变量相区别,后者有天然

的测量值(例如,重量,年龄,收入,被捕次数)。但是,通过对有序类别赋予有序的数值(分数),以定量的方式处理有序变量常常是有益的。

1.1.3 本书的结构

本书的各章可归并为三部分。第一部分由第一章、第二章、第三章组成,叙述1960年以前已出现的分析类别数据的某些标准方法,包括评定两个类别变量的相关性的标准方法以及分析这种关联性如何受第三个变量影响的标准方法。

本书第二部分由第四、五、六、七章组成,介绍类别响应变量的模型。这些模型类似于连续响应变量的回归模型,只是假定响应变量服从二项分布或泊松(Poisson)分布而不是正态分布。我们详细地提出了两类模型:罗吉斯提克(Logistic)模型(也叫罗吉特(Logit)模型)与对数线性(Loglinear)模型。前者应用于响应变量服从二项分布情形,后者应用于响应变量服从泊松分布情形。这两个模型的关系及第四章都表明,这两个模型是一大类模型的特殊情形,通常的关于连续响应变量的正态回归模型也属于这个大类。

本书第三部分由第八章和第九章组成,就两个特殊情形讨论罗吉特模型和对数线性模型。第八章把二项响应模型推广到多项响应模型。多项响应是指响应变量的取值范围是多个类别,而不是两个类别。也考虑这些类别是有序的情形。第九章叙述了成对比较的模型。例如,当我们在两个不同的时间观察同一个个体(或对象)的类别响应时就碰到这样的模型。本书第十章概述了类别数据分析方法的产生历史。

很多分析类别数据的方法需要进行大量的计算。本书附录讨论了SAS和SPSS统计软件的使用法,并就本书中所要执行的计算工作提供了SAS编码。

1.2 抽样模型

类别数据的分析或者任何种类数据的分析,都需要对产生这种数据的随机机制有所假定。对于连续变量数据的回归和方差分析(ANOVA)来说,正态分布起着中心的作用。本节对类别数据讨论两个关键性分布:泊松(Poisson)分布和二项分布。

1.2.1 泊松抽样

在英国,M1是一条主要的南北向的高速公路,有大量的商务车辆和私人车辆使用这条公路。一群不列颠交通研究者研究这条公路穿过田野的路段上致命的汽车事故的发生率。他们记录了一年内这个路段上致命的汽车事故,数据是各周的事故次数。

泊松分布是表示一周里这种致命事故次数的潜在的概率模型,它有一个参数 μ (均值)。设 y 表示一个泊松变量的可能结果。泊松概率的公式是

$$P(y) = e^{-\mu} \mu^y / y!, y = 0, 1, 2, \dots \quad (1.2.1)$$

y 的可能结果均是非负整数。 $y!$ 是 y 的阶乘, $y! = 1 \times 2 \times \cdots \times y, 0! = 1$ 。 e^x 是指数函数, 有时记 e^x 为 $\exp(x)$ 。 e^x 的值可用袖珍计算器计算。指数函数是自然对数的反对数。 $e^a = b$ 等价于 $\log b = a$, 这里 \log 表示自然对数。例如 $e^0 = 1$ 对应 $\log 1 = 0$, 类似地, $e^{0.7} = 2.0$ 对应 $\log 2 = 0.7$ 。本书中所有对数均指自然对数, 底是 $e = 2.718\cdots$ 。

假设在高速公路 M1 上每周致命事故的平均发生率是 2, 对于 $\mu = 2$ 的泊松模型, 一周里致命事故的次数是 0 的概率为

$$P(0) = \frac{e^{-2}2^0}{0!} = e^{-2} = 0.135$$

(注意 $2^0 = 1, 0! = 1$)。类似地, $P(1) = e^{-2}2^1/1! = 2e^{-2} = 0.271$ 。表 1.1 记载了各个 y 值对应的概率, 图 1.1 表示了这些概率。虽然这个分布是对所有的非负整数 y 都有定义, 但 $\mu = 2$ 时所有概率集中在 $y = 0, 1, \cdots, 8$ 的范围内。

表 1.1 均值是 2.0 的泊松分布和二项分布

y	$P(y)$	
	泊松分布	二项分布
0	.135	.107
1	.271	.268
2	.271	.302
3	.180	.201
4	.090	.088
5	.036	.027
6	.012	.005
7	.004	.001
8	.001	.000
9	.000	.000
10	.000	.000
≥ 11	.000	—

注: 二项分布的参数是 $N = 10, \pi = 0.2$

既然每周致命事故的平均次数是 2, 则 t 周内发生致命事故的平均次数是 $2t$ 。例如三周内发生致命事故的次数可用 $\mu = 6$ 的泊松分布来刻画。特别, 三周内无致命事故的概率是 $e^{-6}6^0/0! = e^{-6} = 0.002$ 。这个数等于 $(0.135)^3$, 即三周内的每一周均无致命事故的概率。图 1.1 也描绘了这个均值对应的泊松分布, 这个泊松分布比起 $\mu = 2$ 对应的泊松分布有较大的散布范围。