现代数学基础丛书 147

零过多数据的统计分析及其应用

解锋昌 韦博成 林金官 著

科学出版社

中国科学院科学出版基金资助出版

现代数学基础丛书 147

零过多数据的统计分析及其应用

解锋昌 韦博成 林金官 著

科学出版社

北京

内 容 简 介

本书系统介绍 ZI 数据和相关 ZI 模型的统计推断原理、方法和应用. 内容主要包括：ZI 模型参数的极大似然估计、Bayes 估计、基于经典方法的影响诊断、基于 K-L 距离的 Bayes 影响诊断、ZI 参数和散度参数的假设检验、 ZI 随机效应模型参数的极大似然和 Bayes 估计、基于经典方法的影响诊断、基于 K-L 距离的 Bayes 影响诊断、回归系数和散度参数的假设检验、方差成分检验、ZI 模型及相应的随机效应模型中与均值函数有关的协变量函数形式和联系函数形式的误判检验等.

本书可作为理工科应用统计、公共卫生、生物医学、经济学、生命科学、社会学专业大学生和研究生的教学参考书, 亦可供相关专业的教师、科技人员和统计工作者参考.

图书在版编目(CIP)数据

零过多数据的统计分析及其应用/谢锋昌, 韦博成, 林金官著.
—北京：科学出版社, 2013
(现代数学基础丛书; 147)
ISBN 978-7-03-037283-3

Ⅰ. ①零… Ⅱ. ①谢… ②韦… ③林… Ⅲ. ①统计分析 Ⅳ. ①O212.1

中国版本图书馆 CIP 数据核字(2013) 第 072232 号

责任编辑: 陈玉琢 / 责任校对: 张怡君
责任印制: 钱玉芬 / 封面设计: 陈 敬

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

骏主印刷厂 印刷
科学出版社发行 各地新华书店经销
*
2013 年 4 月第 一 版 开本: B5(720 × 1000)
2013 年 4 月第一次印刷 印张: 14 1/4
字数: 269 000
定价: 58.00 元
(如有印装质量问题, 我社负责调换)

前　言

在公共卫生、生物医学、经济、保险精算、道路安全、制造业和农业等众多领域都存在大量的计数数据. 为了分析这类数据, 常常借助于经典的离散广义线性模型. 然而在实际问题的计数数据中, 往往会含有大量超过标准模型能够预测的取值为零的数据, 称此类数据为零过多 (zero-inflated, 简记为 ZI) 数据, 此时, 标准离散分布可能不再适合分析它们. 取而代之, 近年来兴起的 ZI 模型成为分析零过多数据的有效方法, 受到人们越来越多的重视, 是当今统计学的热点问题之一, 其研究在理论上、应用上都有十分重要的意义.

但是迄今为止, 国内外尚未见到系统介绍这一内容的著作, 本书就是希望填补这方面的空白, 向读者系统介绍零过多数据的统计分析方法及其应用价值. ZI 模型是经典离散模型的推广和发展, 而随着计算机的快速发展和实际数据复杂化程度的提高, 一些适应性更广但比标准离散模型更复杂的模型受到理论和应用工作者越来越多的重视, 诸如负二项模型、双泊松模型、广义泊松模型等. 基于此, 本书首先系统介绍 ZI 数据的基本概念和实际背景以及基本 ZI 模型 (ZI 泊松模型、ZI 二项模型等) 的统计分析方法. 在此基础上, 本书着重介绍更复杂的 ZI 模型的基本理论和实际应用, 其中包括 ZI 负二项模型、ZI 广义线性模型、ZI 广义泊松模型、ZI 双泊松模型等. 其次, 本书系统介绍这些模型的极大似然估计及其 EM 算法、ZI 参数的存在性检验、散度参数和方差成分检验、模型中均值函数的误判检验、全局影响分析和局部影响分析、ZI 数据的 Bayes 统计分析及其马尔可夫链蒙特卡罗 (Markov chain Monte Carlo, MCMC) 算法等问题, 其中也包括作者近年来在零过多数据方面的工作. 本书在详细介绍有关统计理论和方法的同时, 还重点介绍这些理论和方法在公共卫生、生物医学、经济学、保险精算和农业等领域中的具体应用.

本书共分 5 章. 第 1 章引入 ZI 数据的概念, 并基于实际问题介绍一般的 ZI 计数数据和带有重复测量的 ZI 计数数据, 同时为了使读者对 ZI 模型有比较好的了解, 本章还介绍本书涉及的主要分布及其相关性质. 第 2 章介绍 ZI 泊松模型、ZI 二项模型、ZI 负二项模型和 ZI 广义线性模型等几个经典模型的参数估计、ZI 参数的假设检验、统计诊断等问题. 第 3 章研究广义 ZI 泊松回归模型的参数估计、全局影响分析、局部影响分析、ZI 参数和散度参数的存在性检验和齐性检验以及基于累加残差方法的均值函数的误判检验问题. 第 4 章研究广义 ZI 泊松随机效应模型的参数估计, 同时基于最佳线性无偏预测 (BLUP) 型的对数似然函数研究模型的统计诊断、ZI 参数的 score 检验、回归系数和散度参数的存在性检验、散

度参数的齐性检验、基于梯度检验方法的方差成分检验以及基于累加残差方法的均值函数的误判检验等问题. 第 5 章研究广义 ZI 泊松模型及相应混合效应模型的 Bayes 方法, 其中包括 Bayes 估计、Gibbs 抽样、Metropolis-Hastings 算法, 以及基于 Kullback-Leibler 距离的 Bayes 影响分析. 本书每一章都附有较多的应用实例, 并注重介绍数值计算方法和模拟研究的结果.

本书在写作过程中, 自始至终得到科学出版社的关心与帮助, 特别要感谢数理分社的陈玉琢编辑, 她对本书的写作、审定与出版都给予了大力的支持与帮助. 在本书写作过程中, 参考了国内外许多文献, 受益匪浅, 一并对这些文献作者表示衷心的感谢! 同时也要感谢中国科学院科学出版基金、教育部人文社会科学规划基金项目 (11YJA910004)、国家自然科学基金 (11171065, 11271193)、江苏省自然科学基金 (BK2011058) 和江苏省高校自然科学研究计划项目 (11KJB110005) 的资助.

由于作者水平有限, 难免有不妥之处, 恳请同行专家和广大读者提出批评和建议.

作　者

2012 年 7 月于南京

目　录

第 1 章　零过多数据及预备知识

在公共卫生、生物医学、经济学、保险精算、道路安全、保险和农业等众多领域, 都存在大量的计数数据. 为了分析这类数据, 研究者常利用经典的离散分布, 如泊松分布、二项分布或负二项分布等建立模型. 然而, 在实际问题的计数数据中, 往往会出现大量过多取值为零的现象, 如调查人们一天中吸烟的数量时, 其中吸烟 0 支, 即不吸烟的人很多; 研究某药品服用后产生的不良反应的次数时, 其中不良反应次数为 0, 即无不良反应的人很多; 等等. 其中 0 的个数要明显多于泊松、二项或负二项等标准离散分布随机产生的个数, 我们称此现象为零过多 (zero-inflated, ZI) 现象, 此时, 通常的离散分布将不再适合用来刻画它们. 近年来兴起的 ZI 模型成为分析零过多数据的有效方法, 受到人们越来越广泛的重视, 是当今统计学的热点问题之一, 其研究在理论上、应用上都有十分重要的意义. 本书将系统介绍如何对这类数据进行有效的统计推断. 此外为了更加深入有效地研究问题, 计数数据常是通过重复测量得到的 (如果条件允许). 例如, 对若干试验者服用药品后, 每隔一定时间测量他们的不良反应次数, 测量多次, 则这批数据为典型的纵向计数数据. 当然, 这些重复测量数据的研究中也会产生零过多现象, 同时还会产生相关性 (见下一节例 5 和例 6), 本书也将详细介绍如何对这类重复测量的数据进行有效的统计推断. 另外, 随着计算机的快速发展和实际数据复杂化程度的增加, 一些适应性更广但比标准离散模型更复杂的模型受到理论和应用工作者越来越多的重视. 本书首先系统介绍基本的 ZI 模型 (ZI 泊松模型、ZI 二项模型等) 的统计分析方法, 并在此基础上着重介绍更复杂的 ZI 模型的基本理论和实际应用, 其中包括 ZI 负二项模型、ZI 广义线性模型、ZI 广义泊松模型、ZI 双泊松模型等, 这些都是更加有效、更加符合实际的 ZI 模型.

本章介绍 ZI 数据的实际背景及其有关的预备知识. 1.1 节通过两个实例阐述什么是零过多数据; 1.2 节介绍本书涉及不同领域的具体 ZI 数据案例; 1.3 节则介绍常见的经典离散分布. 这些都是本书后面介绍的模型和方法所需要的基础知识.

1.1　什么是零过多数据

近年来, 零过多数据的研究越来越受到理论和应用工作者的重视. 一般情况下, 零过多数据是相对于非零数据而言, 零的个数超过预期出现的数量. 以泊松分布为例, 假定有一组含零很多的计数数据, 其中包含非零数据和部分来自于泊松分布

的零数据, 而余下的零数据则是额外得到的, 这种数据称为零过多数据, 其中额外得到的零在有些文献中也称为结构上的零. 实际上这些额外的零可以看成取值为零的退化总体产生的, 而其余的数据则可认为是非退化总体 (如泊松分布) 产生的. 因此, 零过多数据实际上是退化部分和非退化部分形成的混合分布产生的. 它既可以出现在连续数据中, 也可以出现在离散数据中, 如工业过程中产品的缺陷个数 (Lambert, 1992) (缺陷个数为零的很多)、园艺试验中使用杀虫剂后粉虱的存活数 (Hall and Zhang, 2004) (存活数为零的很多) 等 (参见 1.2 节). 另外, 与零过多数据对应的是零不足数据, 即零的数量比预期产生的数量少, 不过实际问题中这类数据不太常见. 本书主要研究带有零过多的离散数据 (也称为零过多计数数据).

对于零过多计数数据, Gupta 等于 1996 年曾经指出, 当观测到额外的取值为 0 的计数数据时, 如果我们仍用普通的泊松模型进行拟合, 则对于计数数据中取值较小的数据的预测将会产生较大误差. 以下是两个较典型的零过多数据的实例, 它们说明了普通泊松模型拟合时存在的缺陷.

例 1.1.1　HIV 数据(Broek, 1995).

该数据记录了 98 位 HIV 疾病感染者的尿道感染次数, 其频数分布见图 1.1.1. 从图 1.1.1 中可以看出感染 0 次的人特别多, 约占 82.6%, 这是一个典型的零过多数据. 由于是离散型数据, 通常可用泊松分布进行拟合, 其拟合结果见图 1.1.1(a). 但是由图 1.1.1(a) 可知, 其拟合效果很不理想. 图 1.1.1(a) 显示, 拟合预测感染 0 次的期望频数与实际观测频数有较大差距, 而且对于感染 1 次和 2 次的期望频数与观测频数也有很大差距. 因此说明, 应用普通泊松分布拟合 HIV 数据效果不好. 所以 Broek 建议用 ZIP 模型 (即 ZI 泊松模型, 见第 2 章) 拟合 HIV 数据, 其结果列于图 1.1.1(b). 由图 1.1.1(b) 可以看出, 经 ZIP 模型拟合, 由此获得的期望频数与实际观测频数都相当接近, 特别是对于感染 0 次的情形. 这表明用 ZIP 模型拟合 HIV 数据效果得到显著改进. 另外, 两个模型的拟合效果也可以通过 Pearson 拟合优度统计量 χ^2 进一步得到说明. 当 HIV 数据用普通泊松分布拟合时, $\chi^2 = 16.135$, 相应的 p 值为 0.0003, 表明拟合很不好; 而用 ZIP 分布拟合时, $\chi^2 = 1.3723$, 相应的 p 值为 0.2414, 表明拟合优度得到显著改进.

例 1.1.2　Accident 数据(Greenwood and Yule, 1920; Bohning, 1998) (女工事故数据).

该数据记录的是关于军工厂中 647 位女性工人发生事故的次数, 见图 1.1.2, 其中发生 0 次事故的人约占 70%, 这也是一个零过多数据. 与例 1.1.1 类似, 该数据若用泊松分布拟合 (图 1.1.2(a)), 其发生 0 次和 1 次事故的期望频数与实际观测频数差距均较大; 而用 ZIP 分布拟合时 (图 1.1.2(b)), 其差距明显变小. 另外, 用泊松分布拟合时, 其 Pearson 拟合优度统计量的 p 值小于 0.00001, 表明拟合很不好. 若用 ZIP 分布拟合, 则相应的 p 值为 0.0495, 表明拟合优度得到显著改进.

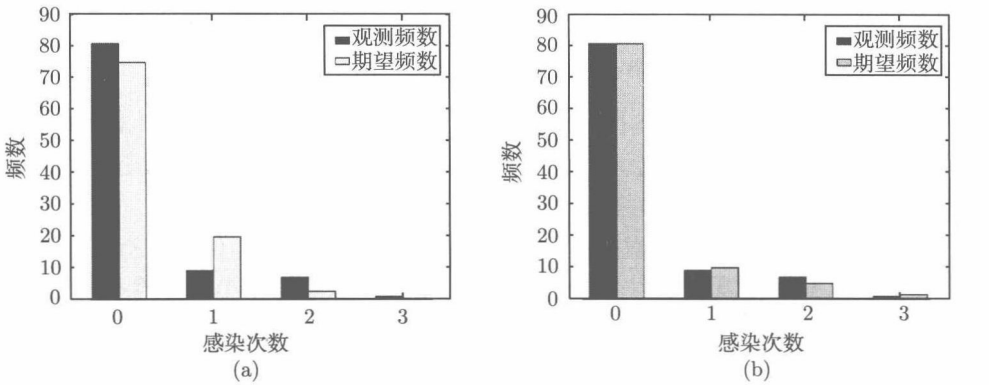

图 1.1.1 尿道感染的观测频数以及泊松和 ZIP 模型预测的期望频数

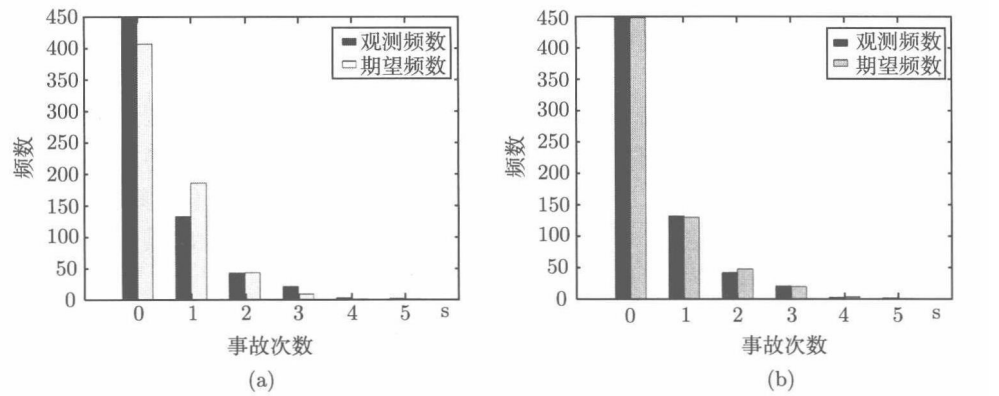

图 1.1.2 事故的观测频数以及泊松和 ZIP 模型预测的期望频数

例 1.1.1 和例 1.1.2 说明对于零过多数据, 由于实际数据与既定的传统模型之间可能存在比较大的偏离, 如果我们不考虑这种偏离, 仍然沿用经典的分析方法, 就可能导致错误的结论. 另外, 在计数数据中除了零过多现象外, 还常常出现偏大离差 (overdispersion) 或偏小离差 (underdispersion) 现象, 若忽视这种现象, 也会导致错误推断 (林金官, 2002). 在实际计数数据中, 零过多现象与偏大或偏小离差现象往往会同时存在, 后面将探讨如何建立适当的模型来同时刻画这两种现象.

1.2 零过多计数数据实际案例

除了 1.1 节介绍的两组零过多计数数据之外, 实际问题中还有很多类似的数据, 下面介绍几个来自不同领域的零过多计数数据, 本书后面各章将经常引用它们, 是

阅读本书有关章节所必需的案例.

1. 机动车保险索赔数据

该保险索赔数据来自于 SAS Enterprise Miner 的数据库, 共有 10303 个原始观测数据, 由于大多数的数据记录不完整, Yip 和 Yau (2005) 仅考虑了最近一年中公司的保单持有人情况, 从而最终获得 2812 个有完整记录的客户.

该数据集包含的信息有索赔概况、保单细节、驾驶记录和保单持有人的详情等. 从索赔概况中可以确定每个投保人的索赔次数; 保单的细节中包含保单编号、客户识别号码、保单的生效日期、家庭或单位所处地区、往返于住处和单位的时间以及投保车辆的价值、种类、用法、颜色; 驾驶记录中包含投保人违规驾驶的记录和在过去七年里有无政府机构吊销保单持有人的驾照; 个人资料中包含性别、年龄、出生日期、婚姻状况、子女数目、每年的收入、工作类别和教育水平等情况. 在以上的信息中, 我们可以发现保单的细节、驾驶记录和个人资料可能包含着影响索赔经历的潜在风险因素.

我们知道, 索赔次数是合理确定保费的一个重要依据, 因此, 合理刻画索赔次数对车险业务保费的计算具有现实的意义. 这里的索赔次数从 0 到 5 不等, 其中一个有趣的特点是零索赔的比例非常高, 大约 60.7%的投保人没有索赔, 12.5%的保单持有人提出一项索赔, 平均索赔频率约为 0.82. 关于索赔次数的具体情况见图 1.2.1, 可以看出零次索赔的图形明显高于其他情形.

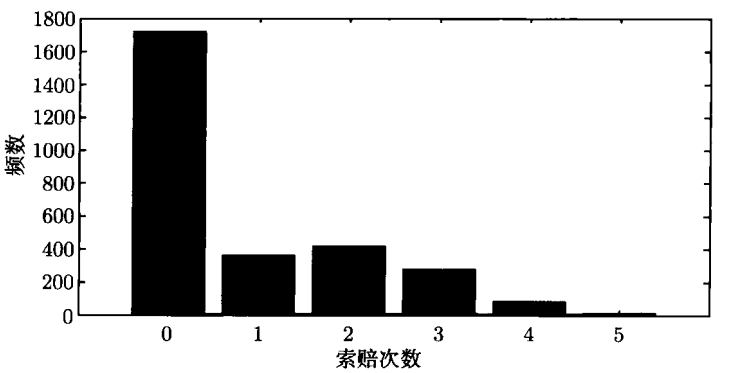

图 1.2.1　索赔次数的观测频数

显然, 这是一个零过多数据, 对于该数据, 第 2 章将利用 ZIP 模型以及零过多负二项 (ZINB) 模型进行统计分析, 从而可以找出对索赔次数有较大影响的风险因素.

2. 医院门诊数据

为了全面了解人们如何使用和支付医疗卫生服务, 美国在 1987 年和 1988 年进

行了医疗支出统计调查 (NMES), 共涉及全美 15000 个家庭约超过 38000 个人. 该调查随机采访了家庭健康保险覆盖面、涉及的服务以及支付这些服务的成本和资源等. 数据集中除了医疗保健数据外, NMES 还提供了卫生状况、就业情况、社会人口特征、经济状况等信息.

在本书中, 我们仅考虑中西部地区 66 岁及以上的男性且享受私人医疗保险的子样本, 共 401 个观测值. 研究的数据中涉及的指标主要有医院门诊次数、慢性病数 (癌症、心脏病、胆囊问题、肺气肿、关节炎、糖尿病等)、日常生活活动限制情况 (若有则为 1)、年龄 (年)、种族 (若该人是非洲裔美国人则取 1)、婚姻状况 (若结婚则为 1)、学校受教育年限、家庭收入、就业情况 (若有工作则为 1) 等. 关于数据的详细说明可参见 Deb 和 Trivedi (1997). 我们发现这组数据中 0 很多, 占总数的 73.57%, 且从图 1.2.2 中也可看出, 零次门诊的频数图明显高于其他情形. 因此, 这是一个零过多数据, 为了进一步分析该数据, 第 3 章和第 5 章将分别利用零过多广义泊松模型和零过多双泊松模型对其进行统计分析.

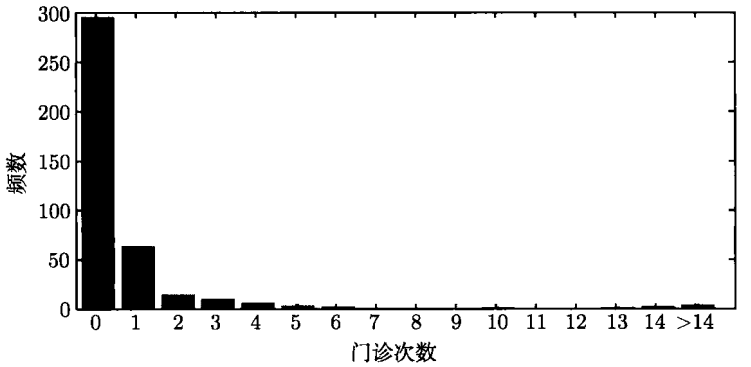

图 1.2.2 医院门诊次数的观测频数

3. *旅游数据*

该数据来自 Cameron 和 Trivedi (1998), 共有 659 个观测值, 目的是探讨对 1980 年划船到东德克萨斯州 Somerville 湖的旅游次数的影响因素. 主要涉及 7 个指标: 简明的个人品质等级、划水的体验应答、收入、消费哑变量 (若每年消费者在 Somerville 湖有消费, 则为 1, 否则为 0)、旅游到 Conroe 湖的消费、旅游到 Somerville 湖的消费、旅游到 Houston 湖的消费. 关于到东德克萨斯州 Somerville 湖的旅游次数具体情况见图 1.2.3, 明显可以看出没去过的人数显著高于去过的, 实际上没去过旅游的人数占总调查人数约 63.3%. 显然, 这是一个零过多数据, 读者可以在第 3 章看到, 我们将利用零过多广义泊松模型和零过多双泊松模型对其进行进一步的统计分析.

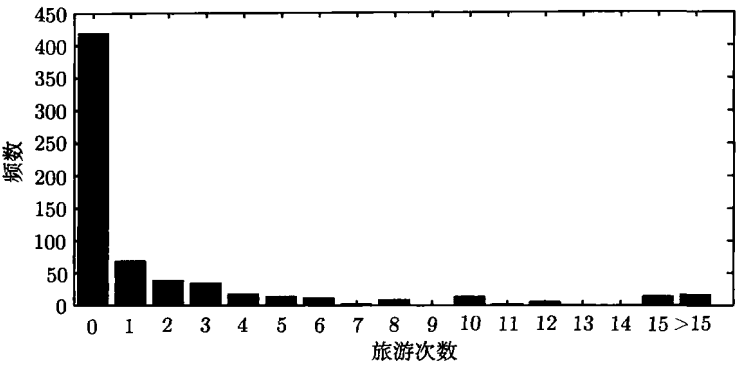

图 1.2.3　旅游次数的观测频数

4. 室性早搏数据

该数据最初由 Berry (1987) 以计数数据形式给出, 后来, Farewell 和 Sprott (1988) 将其作为比例数据对其进行分析. 该数据涉及 12 位患者, 他们患有频繁的室性早搏 (PVC), 为此, 他们服用了抗心律失常药物. 分别在用药前后 1min 记录了 12 位患者的心电图, 同时还记录了 PVC 次数. 具体数据见表 1.2.1 (Berry, 1987), 来自于 Deng 和 Paul (2000) 的表 1.2.1.

表 1.2.1　12 位患者的 PVC 数据

患者编号	用药前 PVC 次数 (x_i)	用药后 PVC 次数 (y_i)	总的 PVC 次数 (m_i)
1	6	5	11
2	9	2	11
3	17	0	17
4	22	0	22
5	7	2	9
6	5	1	6
7	5	0	5
8	14	0	14
9	9	0	9
10	7	0	7
11	9	13	22
12	51	0	51

从表 1.2.1 中可以看出, 用药后患者未出现室性早搏的占 58.33%. 另外, 从图 1.2.4 也可以看出, 出现 0 次室性早搏的患者数明显高于有室性早搏的患者数. 显然, 这是一个零过多数据, 在第 2 章, 我们将利用零过多二项分布 (ZIB) 模型对其进行统计分析.

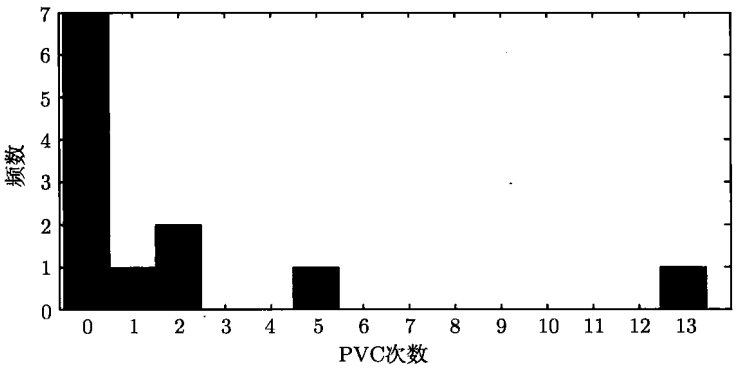

图 1.2.4 用药后 PVC 的观测频数

5. *粉虱数据*

该数据来自于利用杀虫剂控制温室栽培的一品红上的银叶粉虱的试验 (Hall and Zhang, 2004). 试验设计是完全随机分组的, 每周重复测量, 共计 12 周. 试验中每三株一品红作为一个试验单位, 共有 18 个试验单位, 它们被随机分成三个不同的区组进行 6 种不同的试验. 当粉虱出现于固定在叶子上的笼子里两天后, 开始按周计量其中存活的昆虫数, 共测量 12 周, 最终得到 640 个观测数据, 其中存活 0 只昆虫的情形大约占 53%, 具体情况见图 1.2.5.

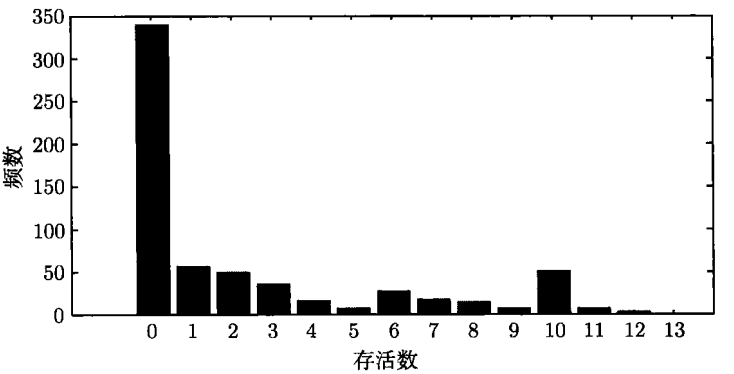

图 1.2.5 存活昆虫数的观测频数

显然, 这是一个零过多数据, 而且是带有重复测量的零过多数据, 流行的分析方法是采用随机效应模型进行刻画, 第 4 章将基于零过多广义 ZI 泊松随机效应模型进行详细研究.

6. *制药数据*

该数据来自于某制药公司 (Min and Agresti, 2005), 目的是研究利用两种不同

方案治疗特殊疾病时产生的副作用次数. 数据中共涉及 118 位患者, 其中 59 人随机安排接受方案 A (TRT1) 治疗, 另外 59 人则接受方案 B(TRT2) 治疗. 然后, 对患者进行 6 次随访, 每次计量产生的副作用次数, 由于副作用次数随着随访时间间隔而有所变化, 我们将其作为协变量 (定义为 Time) 引入到模型中. 该数据中大约有 83%的观测为零次副作用, 具体见图 1.2.6, 从中可发现产生零次副作用的图形最高. 显然, 这也是一个零过多数据, 该数据与粉虱数据属于同一类型, 是带有重复测量的零过多数据. 为了探究副作用产生的机制, 我们在第 4 章和第 5 章基于零过多广义泊松随机效应模型对其进行较详细的统计分析.

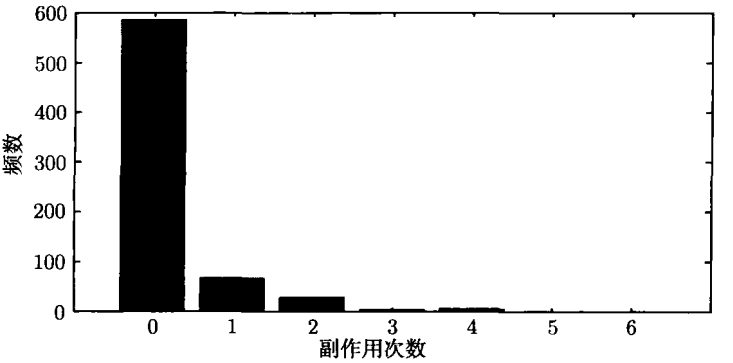

图 1.2.6　副作用次数的观测频数

1.3　预备知识 —— 常用的离散分布

本书后面讨论的模型主要产生于若干常用的离散分布, 下面对所涉及的分布作简要介绍.

1. *泊松分布*

在实际离散数据分析中, 泊松分布是最基本、最常用的模型. 现在假定随机变量 Y 服从泊松分布, 则其概率函数可以写成如下形式:

$$P(Y=y)=\frac{\lambda^y}{y!}\mathrm{e}^{-\lambda},\quad \lambda>0,\ y=0,1,2,\cdots. \tag{1.3.1}$$

根据式 (1.3.1) 易得泊松分布的期望和方差为

$$E(Y)=\mathrm{Var}(Y)=\lambda.$$

方差和期望相等是泊松分布的一个重要特征, 也称为等偏差 (equidispersion), 它在后面的研究中起着关键作用, 当方差和期望不等时就产生了偏大离差 (方差大于期望) 或偏小离差 (方差小于期望) 的情形.

2. 二项分布

当计数数据有界时, 我们常考虑利用二项分布进行刻画. 假定随机变量 Y 表示 m 次试验中事件成功的次数, 即其服从二项分布, 且概率函数为

$$P(Y=y)=\frac{m!}{y!(m-y)!}\pi^y(1-\pi)^{m-y},\quad y=0,1,2,\cdots,m. \tag{1.3.2}$$

此时易得 Y 的期望和方差分别为

$$E(Y)=m\pi,\qquad \mathrm{Var}(Y)=E(Y)(1-\pi).$$

3. 负二项分布

当计数数据中方差与期望不等时, 我们常利用负二项分布取代泊松分布进行建模. 假定随机变量 Y 服从负二项分布, 且具有下面概率函数

$$P(Y=y)=\frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\Gamma(y+1)}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{y},\quad y=0,1,2,\cdots, \tag{1.3.3}$$

其中 $\alpha\geqslant 0$, $\theta\geqslant 0$, $\Gamma(\cdot)$ 表示 gamma 函数 $\Gamma(s)=\displaystyle\int_0^{+\infty}z^{s-1}\mathrm{e}^{-z}\mathrm{d}z$ $(s>0)$. 记 $Y\sim NB(\alpha,\theta)$.

此时, 负二项分布的期望和方差分别为

$$E(Y)=\alpha\theta \tag{1.3.4}$$

和

$$\mathrm{Var}(Y)=\alpha\theta(1+\theta)=E(Y)(1+\theta). \tag{1.3.5}$$

由于 $\theta\geqslant 0$, 所以负二项分布的方差大于其期望, 即存在偏大离差现象, 而且当 $\theta\to 0$ 时, 偏大离差现象将消失.

负二项分布有多种参数形式, 为了能够利用该分布建立回归模型, 可以对其进行期望参数化, 即

$$\lambda=\alpha\theta, \tag{1.3.6}$$

其中 λ 是该分布的期望. 根据式 (1.3.6) 有以下两种常用的参数化形式.

(1) $\alpha=\lambda/\theta$, 此时方差具有下面形式:

$$\mathrm{Var}(Y)=\lambda(1+\theta), \tag{1.3.7}$$

易见方差是期望的线性函数, Cameron 和 Trivedi (1986) 称其为 I 型负二项分布, 记为 NBI, 其对应的概率函数为

$$P(Y=y)=\frac{\Gamma(\lambda/\theta+y)}{\Gamma(\lambda/\theta)\Gamma(y+1)}\left(\frac{1}{1+\theta}\right)^{\frac{\lambda}{\theta}}\left(\frac{\theta}{1+\theta}\right)^{y}. \tag{1.3.8}$$