

医学统计学基础 与典型错误辨析

主编 胡良平 李子建

军事医学科学出版社

医学统计学基础与典型错误辨析

主 编 胡良平 李子建

编 委

李子建 刘惠刚

童中彪 胡良平

军事医学科学出版社

2003·北京

内容提要

本书较全面地揭示了统计学应用中存在的问题,提出了切实可行的解决问题的对策和建议。详细地介绍了学习统计学的策略、所必需的基本知识、常用的描述性统计分析方法和假设检验方法。这些方法包括“资料类型、相对数、分位数、平均指标、变异指标、平均与变异指标的结合应用、统计表与统计图、概率与概率分布的概念、常用离散型随机变量概率分布、常用连续型随机变量概率分布、假设检验及其应用、统计分析方法合理选用、常用定量和定性资料统计分析方法、相关回归分析方法、资料综合分析方法”。本书还从2 000多篇文献中精选出百余篇文献,对其中误用和滥用统计学的典型案例逐一作了精辟地分析和释疑,这些内容将有助于人们减少统计学应用中常犯的错误,增强人们正确应用统计分析方法解决实际科研问题的能力。

本书可作为学习常用统计分析方法的参考书,可作为辨析误用和滥用统计学的“透视镜”,更是学习统计学、讲授统计学和巧用统计学的方法学用书。

图书在版编目(CIP)数据

医学统计学基础与典型错误辨析/胡良平,李子建主编. - 北京: 军事医学科学出版社, 2003. 7

ISBN 7-80121-520-6

I. 医… II. ①胡… ②李… III. 医学统计-参考资料 IV. R195.1

中国版本图书馆CIP数据核字(2003)第042358号

出 版: 军事医学科学出版社出版

地 址: 北京市太平路27号

邮 编: 100850

联系电话: 发行部: (010) 66931034

66931048

编辑部: (010) 66931127

传 真: (010) 68186077

E-MAIL: mmsped@nic.bmi.ac.cn

印 刷: 春园印装厂

装 订: 春园印装厂

发 行: 新华书店总店北京发行所

开 本: 787 mm×1092 mm 1/16

印 张: 24.25

字 数: 604千字

版 次: 2003年9月第1版

印 次: 2003年9月第1次

印 数: 1-3500册

定 价: 36.00元

(购买本社图书,凡有缺、损、倒、脱页者,本社发行部负责调换)

前言

很多人经常提出这样的问题：“为什么我们学了多遍统计学仍不得要领，每到用时就出错？”道理很简单，虽然学了多遍统计学，但每遍学的内容都大同小异，学的内容少，真正掌握的内容就更少了，而要解决的问题却远远超出了自己所掌握的知识范围；每次几乎都是脱离实际地学统计，不会将实际问题转化为统计问题，用十分有限的“几个招术”去对付各种各样复杂的统计问题，生搬硬套，张冠李戴，出错就成为难以避免的事了。事实正是如此，在我国的许多科研成果档案材料中，在许多公开发表的科技期刊的学术论文中，甚至在一些电视和报纸等新闻媒体中，误用和滥用统计学的现象相当普遍，以至于严重影响了我国科技工作的科学性与严谨性，在某种程度上阻碍了科技事业的创新和发展。人类进入 21 世纪，我国已加入 WTO，我国的科技事业要赶上世界发达国家的先进水平，必需在许多重要方面符合国际惯例，与世界接轨。这是人心所向，形势所迫！

笔者从事生物医学统计学工作已有 20 年历史，接触并解决了来自医学科研、临床试验和医学期刊编辑等不同领域里的大量统计学方面的问题，积累了解决各类统计问题的实践经验和丰富资料，主办和主讲“全国生物医学统计学和 SAS 软件培训班”10 期，产生了良好的社会效益。为了使这些宝贵的经验能更好地为我国的改革和经济建设服务，为了更好地帮助广大生物医学科技工作者正确地运用统计学，解决相关领域里的各种统计学问题，笔者认为加强医学统计学的基础是十分必要的，同时，针对医学期刊论文中误用统计学的典型案例进行辨析，使人们“吃一堑长一智”，这对人们彻底纠正“学了多遍统计学仍不会正确应用”的现象是极为有利的。值得庆幸的是：本室即将毕业的研究生李子建在 3 年学习中，刻苦钻研，阅读了发表在 150 多种医学杂志上的 2 000 多篇医学文献，经过精心策划和筛选，找出了数以百计的误用和滥用统计学的典型案例，并逐一作了分析和释疑，确凿的证据说明相当多的原文的结论是错误的。这些精彩案例和深入浅出地分析讲解，无论对于医学统计学教师、广大医学生，还是以应用为主的广大医学科研工作者和医学期刊编辑，都将是不可多得的生动教材，对提高我国医学统计学的教学质量和医学统计学的应用水平，将起到不可估量的作用。

医学统计学的地位和作用：医学统计学是将概率论和数理统计的原理和方法应用于医学科研的一门应用性学科，用于帮助医学科研工作者透过许多偶然现象，分析和揭示事物的内在规律。

20 世纪 40 年代末，《英国医学杂志 (BMJ)》发表了用随机化实验方法评价链霉素治疗肺结核的临床报告，由于采用了随机对照的临床试验，较客观地肯定了链霉素对结核病的疗效，并基本上否定了以往作为定论的“金化合物”等所谓特效药的疗效。此后，医学统计学在医学科研中的应用逐渐得到普及，成为医学科研工作中不可缺少的分析和解决问题的工具，极大地促进了医学科学的发展。现在，几乎所有的医学期刊都要求对医学论文中报告的实验或观察数据进行统计学处理，以保证医学科研工作的科学性和严谨性。《美国医学会杂志 (JAMA)》社的一位主编更是将统计学比喻为“现代医学大厦的一个重要支柱”，用以说明医学统计学在医学领域中的重要地位和作用。严格地说，将统计学的作用仅局限于统计描述和统计分析方面是很不明智的。无论用多么高超的统计分析技术来处理歪曲事实的统计资料，

必将得出荒谬的结论。也就是说,统计学的作用更主要的是体现在“统计研究设计”上。正确制定并严格实施统计研究设计方案,是充分发挥统计学作用的最有力的证据。

医学统计学的应用状况:医学统计学的理论基础是概率论和数理统计,其应用情况直接影响医学科研工作的质量,应用正确,能保证科研工作的顺利进行,使研究结果更具有说服力。反之,如果使用不当或误用、滥用,就会缺乏说服力和可靠性,有时会招摇撞骗,以假乱真,甚至会导致得出错误的结论。

国内外统计学工作者对医学统计学的应用情况进行过多次调查,结果不容乐观。凌瑞珠曾系统地阅读了《中华医学检验杂志》,发现其原著中50%~70%存在统计误用,并认为国内其他医学杂志的统计误用也多处于类似水平。王倩等调查了1995年国内5种具有较高学术水平、发行量较大的中华系列杂志发表的论文,发现54%的文章不能够正确应用统计分析方法。程萍等评阅了《中华内科杂志》1993~1995年论著栏内发表的论文,68%存在统计学方法上的问题或错误。毛宗福等报告了5种中华核心期刊论著中统计推断应用误用率为62%。国外学者对一些英文权威医学期刊调查表明,统计学误用率为50%的不在少数,有的甚至高达70%或100%。Glantz调查了1977年《Circulation Research》和《Circulation》杂志中发表的文章,在考虑使用统计学方法的文章中,分别有61%和44%具有统计学错误。Kanter等调查了1992年7月到1993年6月《Transfusion》杂志发表的文章,发现75%的文章至少存在一处统计假设检验、统计分析和结果解释方面的错误。一些专家认为,近30年来,临床论著中统计应用缺陷没有发生根本改善,误用率仍居高不下。

很多医学实际科研工作者对统计学的作用重视不够,突出表现在忽视医学科研设计,在统计分析时盲目套用统计分析方法,对统计分析结果解释时轻描淡写,一笔带过。把统计学当作无关紧要的“修饰物”,严重影响了医学科研工作的科学性与严谨性。由于轻视或误用统计学而导致得出错误结论的例子并不鲜见。如1978年8月9号光明日报刊登一条科技信息:研究者对美国20个城市做饮水氟化研究,10个城市的饮水氟化,而另10个城市未氟化作对照。结论认为饮水氟化有致癌作用。但过了一个时期,光明日报对此消息进行了更正。原来前述论文发表后,受到美国癌肿协会和英国皇家统计协会的怀疑,他们派人对此进行了调查,发现研究者只是对此资料进行了单因素分析,并没有对种族、生活环境等混杂因素进行校正,对该批数据重新进行多因素统计分析,在扣除了其他混杂因素的影响后,其结论是:饮水氟化没有发现有致癌作用,相反,却略有保护作用。可见研究者的统计学知识和分析策略对保障科研工作的科学性与严谨性具有重要作用。

传统医学统计教学方法存在的不足:传统的医学统计教学模式往往是“填鸭灌输式”或“知识继承型”的教学方法,一般是教师在上面讲,学生在下面听,忙于记笔记、课后做习题。这种教学模式所产生的效果并不理想,医学统计学这门课具有其自身特点:一方面理论比较抽象,概念多,公式繁杂,学生大部分来自生物医学和临床医学专业,缺乏概率论和数理统计方面的背景知识,往往对医学统计学中的概念含义不能正确理解,特别是对概念之间的联系及区别难以掌握,只得去死记硬背,应付考试,学习兴趣普遍不高。另一方面,医学统计学实践性特别强,传统的教学模式在单纯进行统计学考试时,一般显现不出其弊端来,这是因为课本上的经典例题、习题和考题都是加工提炼后的标准型,学生一般不难发现其对应的方法。然而,医学统计学不单单是一门课程,更是一门应用性很强的,解决实际科研问题的学问,当使用者从事实际科研工作时,其所面对的不是例题那样的标准型,而是实际问题

的原型，他们缺乏处理实际问题的经验和能力，往往感到比较复杂，无从下手。很多人在进行统计分析时或是参照教科书生搬硬套，或是根据相关文章所采用的统计分析方法进行简单模仿，不能将实际问题正确地转化为统计学问题，不能根据资料的设计类型、性质和分析目的灵活选用合适的统计分析方法。事实上很多参考文献其统计学应用本身就存在严重问题，这样以讹传讹，在一定程度上助长了统计学的滥用。

随着现代统计学和计算机技术的迅猛发展，一些新的统计学方法和技术逐渐成熟并得到广泛应用，统计软件的功能也日益强大，并促使医学研究向深度和广度发展，传统医学统计学教学模式的“单向性、机械性”等方面的不足更加凸显出来。如何使学习者掌握正确的统计学概念，形成正确的统计思想，并运用学到的统计学知识解决实际问题，达到学以致用目的，这些都对医学统计教学提出了新的更高的要求。

撰写本书的目的和意义：能否把各种实际问题正确地转化为统计问题，能否合理选用统计学方法，并且正确运用统计学的理论和方法进行分析和处理，解决实际问题，是学好医学统计学的难点所在。也是衡量医学统计学教学质量的“金标准”。

针对当前医学统计学误用、滥用现象严重的现状，尤其表现在医学期刊中统计学错误较多的现状，本书探索有别于传统医学统计学的新的教学模式。采用逆向法，从公开发表的医学论文中的各种错误实例入手，给出问题的原型，使读者“虚拟地”置身于科研第一线，去感受一个个活生生的科研实例，通过分析对问题有比较全面、准确的理解并找出解决问题的基本途径和方法，通过这种类似“实战”的模式，使读者深刻理解、熟练掌握统计学基本概念和方法，并能达到举一反三、融会贯通之效果。从而可减少人们在实际应用统计学中出现错误的机会，增强处理各种实际问题的能力。

材料与方法：广泛查阅近十年来在公开刊物上发表的医学论文，涉及医药期刊约 150 种，范读和精读的论文约 2 000 篇。分析作者在文章中所表现出的统计学方面存在的主要问题与错误，并做好登记与归纳，积累素材。按医学统计学的知识结构与理论体系进行内容安排，分为实验设计、资料的表达与描述、定量资料的统计分析、定性资料的统计分析、相关与回归分析、结果的解释、多元统计及资料的综合分析等几部分，从各章中的知识点和容易混淆的概念出发，在众多实例中筛选出存在代表性错误的文章，对其进行必要的资料整理，首先给出问题的原型，陈述原作者的研究目的、实验设计、所选取的统计分析方法及主要结论，启发读者发现错误；然后对其进行分析和释疑，并对相关理论进行讲解和辨析，给出正确的解决方案，引导读者建立正确的统计分析思想。

本书围绕“如何巧学统计学、如何巧授统计学和如何巧用统计学”的主题展开，以“加强医学统计学基础和辨析误用统计学的典型案例”为切入点，以人们“学、用统计学的重点和难点”为突破口。本书由 3 篇组成：**第 1 篇医学统计学学习策略**，介绍了医学科研和医学写作中存在的统计学问题及其对策，结合在多年统计教学和咨询中总结出来的经验和体会，详细地介绍了巧学、巧授和巧用医学统计学的方法与技巧；**第 2 篇医学统计学基础**，系统地介绍了学、用统计学的必备基础知识和常用统计分析方法；**第 3 篇医学统计典型错误辨析**，详细列举了医学期刊论著中经常出现的各类误用和滥用统计学的典型案例，并从统计学和专业角度作了全面的分析和释疑，体现了统计学不仅是一门应用科学，而且是一门艺术之魅力。

最后，在本书即将问世之际，笔者十分真诚地感谢首都医科大学继续教育学院刘惠刚副教授给予的大力支持和帮助，感谢研究生李子建、董中彪、李峰青的鼎力协助。

为了方便广大读者学习和使用本书，笔者将书中的全部 SAS 引导程序放在生物医学统计
咨询网站上供读者下载，网址：<http://www.stathlp.com>。

由于笔者水平有限，缺点和错误在所难免，恳请广大读者批评指正！

胡良平 李子建

2003 年 5 月 1 日于北京

目 录

第 1 篇 医学统计学学习策略

第 1 章 医学科研与论文中存在的统计学问题及对策 1

- 1.1 统计学与科技事业的关系 1
 - 1.1.1 统计学与科学研究的关系 1
 - 1.1.2 统计学与科研管理的关系 2
 - 1.1.3 统计学与科研成果的关系 2
 - 1.1.4 统计学与科技期刊的关系 2
- 1.2 医学论文中统计学错误的概述 3
 - 1.2.1 引言 3
 - 1.2.2 医学期刊论文中令人担忧的问题 3
 - 1.2.3 医学期刊论文中统计学错误的种类 4
 - 1.2.4 统计错误产生和延续的原因 5
 - 1.2.5 努力纠正统计错误的必要性 6
- 1.3 纠正医学期刊论文中统计学错误的建议与对策 6
 - 1.3.1 纠正统计错误的可行性 6
 - 1.3.2 纠正统计错误的建议 7
 - 1.3.3 解决疑难统计问题的对策 8
- 1.4 影响我国医药科技事业发展的要因分析及其对策研究 8
 - 1.4.1 存在的问题及其症结 8
 - 1.4.2 解决问题的对策与建议 10
- 1.5 逆向统计教学法的实践与认识 11
 - 1.5.1 为什么很多人学了多遍统计学仍不得要领 11
 - 1.5.2 逆向统计教学法的产生背景 12
 - 1.5.3 逆向统计教学法的具体实施 12
 - 1.5.4 逆向统计教学法的作用与效果 13

第 2 章 统计学的巧授巧学巧用方法 15

- 2.1 统计学的巧授方法 15
 - 2.1.1 总的指导思想 15
 - 2.1.2 统计学绪论的讲授方法 15
 - 2.1.3 统计研究设计的讲授方法 16

2.1.4 统计资料描述与表达的讲授方法.....	17
2.1.5 定量资料统计分析的讲授方法.....	18
2.1.6 定性资料统计分析的讲授方法.....	18
2.1.7 相关与回归分析的讲授方法.....	19
2.1.8 资料综合分析的讲授方法.....	19
2.2 统计学的巧学方法.....	20
2.2.1 总的指导思想.....	20
2.2.2 统计学绪论的学习方法.....	21
2.2.3 统计研究设计的学习方法.....	22
2.2.4 统计资料描述与表达的学习方法.....	22
2.2.5 定量资料统计分析的学习方法.....	22
2.2.6 定性资料统计分析的学习方法.....	23
2.2.7 相关与回归分析的学习方法.....	23
2.2.8 资料综合分析的学习方法.....	24
2.3 统计学的巧用方法.....	24
2.3.1 总的指导思想.....	24
2.3.2 巧用实验设计方法的基本思想.....	25
2.3.3 巧用定量资料统计分析方法的基本思想.....	25
2.3.4 巧用定性资料统计分析方法的基本思想.....	25
2.3.5 巧用相关与回归分析方法的基本思想.....	25
2.3.6 巧用统计资料综合分析方法的基本思想.....	26
2.4 演示“统计学绪论”的讲授方法.....	26
2.4.1 讲授计划.....	26
2.4.2 先问学生几个问题.....	26
2.4.3 统计学内容概述.....	27
2.4.4 医学统计学运用的现状.....	28
2.4.5 典型案例精彩回放.....	28
2.4.6 树立信心、寻找对策.....	33
2.4.7 与学好统计学有关的问题.....	34
2.4.8 人们常感困惑的问题.....	36

第2篇 医学统计学基础

第3章 描述性统计方法.....39

3.1 资料类型.....	39
3.1.1 资料类型的两种划分方法.....	39
3.1.2 资料类型识别中常犯的错误.....	41

3.2 相对数	42
3.2.1 相对数的种类与计算	42
3.2.2 相对数的应用	43
3.3 分位数	48
3.3.1 分位数的种类与计算	48
3.3.2 分位数的应用场合	49
3.3.3 分位数的应用举例	49
3.4 平均指标	51
3.4.1 平均指标的种类与计算	51
3.4.2 平均指标的应用场合	53
3.4.3 平均指标的应用举例	56
3.5 变异指标	59
3.5.1 变异指标的种类与计算	59
3.5.2 变异指标的应用场合	62
3.5.3 变异指标的应用举例	62
3.6 平均指标与变异指标结合应用	64
3.6.1 平均指标与变异指标应用中常见的错误	64
3.6.2 平均指标与变异指标的正确应用	66
第 4 章 统计表与统计图	70
4.1 统计表及其应用	70
4.1.1 统计表的种类与编制要领	70
4.1.2 统计表的正确应用	71
4.2 统计图及其应用	76
4.2.1 统计图的种类与绘制要领	76
4.2.2 统计图的正确应用	76
第 5 章 概率与概率分布	95
5.1 概率与概率分布的概念	95
5.1.1 概率的定义	95
5.1.2 概率的分布与期望值	98
5.1.3 Bayes 理论与主观概率	101
5.2 常用离散型随机变量概率分布	103
5.2.1 二项分布	103
5.2.2 Poisson 分布	110
5.3 常用连续型随机变量概率分布	115
5.3.1 正态分布	115

5.3.2 χ^2 分布	119
5.3.3 t 分布	122
5.3.4 F 分布	126
第 6 章 假设检验及其合理选用策略	130
6.1 假设检验及其应用	130
6.1.1 引言	130
6.1.2 假设检验的基本概念	130
6.1.3 假设检验方法的分类	132
6.1.4 部分假设检验方法的应用	132
6.1.5 常见变量变换方法简介	143
6.2 统计分析方法合理选用策略	144
6.2.1 概述	144
6.2.2 科研工作流程图	145
6.2.3 统计研究设计	145
6.2.4 资料收集、整理、表达和描述	145
6.2.5 假设检验合理选用策略	145
第 7 章 常用定量资料统计分析	148
7.1 定量资料统计分析概述	148
7.1.1 何为定量资料的统计分析	148
7.1.2 定量资料统计分析常见错误及其不良后果	148
7.1.3 合理选用定量资料统计分析方法的关键	148
7.2 正确辨析实验设计类型的 3 种方法	149
7.2.1 平衡的多因素实验设计资料的“对号入座”辨析法	149
7.2.2 平衡的多因素实验设计资料的“结构变形”辨析法	150
7.2.3 非平衡的多因素实验设计资料的“拆分”辨析法	152
7.3 适合选用 t 检验的 3 种实验设计类型及其实现	153
7.3.1 单组设计定量资料的 t 检验	153
7.3.2 配对设计定量资料的 t 检验	157
7.3.3 成组设计定量资料的 t 检验	160
7.4 无法考察交互作用的几种常见实验设计类型及其定量资料的方差分析	164
7.4.1 对定量资料进行方差分析的前提条件	164
7.4.2 方差分析的基本思想	164
7.4.3 几种最常见的实验设计类型定量资料的方差分析	165
7.5 可以考察交互作用的两种常见实验设计类型及其定量资料的方差分析	181
7.5.1 关于实验因素之间交互作用的考虑	181

7.5.2 析因设计应用举例	181
7.5.3 正交设计应用举例	185
第 8 章 常用定性资料的统计分析	189
8.1 定性资料统计分析方法概述	189
8.1.1 何为定性资料的统计分析	189
8.1.2 正确处理定性资料的关键技术	189
8.1.3 列联表资料的分类及对应的统计分析方法	190
8.2 定性资料统计分析应用举例	198
8.2.1 四类 2×2 列联表资料统计分析举例	198
8.2.2 四类 $R \times C$ 列联表资料统计分析举例	204
第 9 章 直线相关与回归分析	218
9.1 相关与回归分析方法概述	218
9.1.1 与相关分析有关的概念	218
9.1.2 与回归分析有关的概念	219
9.1.3 直线相关与回归分析的异同点	221
9.1.4 直线相关与回归分析的关键点	222
9.2 直线相关与回归分析的计算与应用	225
9.2.1 相关系数 r 的计算与假设检验	225
9.2.2 截距 a 和斜率 b 的计算与假设检验	226
9.2.3 直线相关与回归分析的应用	229
第 10 章 资料的综合分析	239
10.1 资料综合分析方法概述	239
10.2 资料综合分析应用举例	239
第 3 篇 医学统计典型错误辨析	
第 11 章 实验设计中存在的统计学错误	242
11.1 实验设计没有遵循随机原则	242
11.2 以随意代替随机的例子	242
11.3 重复设置对照组	243
11.4 对照不全	244
11.5 缺乏对照组	245
11.6 对照组的设立与实验设计的目的不符	246
11.7 样本含量太小	247

11.8 仅测量一个样品就作出结论.....	247
11.9 忽视混杂因素的干扰	248
11.10 忽视至关重要的非实验因素.....	248
11.11 各组之间均衡性不好.....	249
11.12 实验分组不合理	249
11.13 缺乏完善的实验设计方案.....	250
11.14 与中心思想不相干的实验设计.....	252
11.15 用单因素设计取代多因素设计.....	252
11.16 实验因素的各水平组合不完全.....	254
11.17 适合采用裂区设计的实验研究问题.....	256
11.18 缺乏与设计类型相匹配的收集实验数据的表格.....	257
11.19 没有按照实验设计类型的要求进行实验和收集数据.....	258
第 12 章 资料的表达与描述中存在的统计学错误	260
12.1 误用正态分布法描述呈偏态分布的资料.....	260
12.2 以“比”代“率”	260
12.3 以“率”代“比”.....	261
12.4 计算相对数时分母太小	261
12.5 数据造假	262
12.6 统计表纵、横标目颠倒	263
12.7 统计表中数据含义不清	264
12.8 同一表中表达不同的内容	266
12.9 用一个复式条图表达多个指标	269
12.10 用复式条图表达适合用线图表达的资料.....	271
12.11 用线图表达适合用复式条图表达的资料.....	272
12.12 统计图坐标轴上的刻度不符合数学原则.....	273
第 13 章 定量资料分析中存在的统计学错误	276
13.1 误将定量资料判为定性资料从而误用 χ^2 检验.....	276
13.2 用 χ^2 检验分析定量资料	277
13.3 误用 χ^2 检验处理具有一个重复测量的两因素设计定量资料.....	277
13.4 忽视 t 检验、方差分析的前提条件	278
13.5 误用成组设计资料的 t 检验分析配对设计的资料.....	279
13.6 误用一般 t 检验分析配伍组设计的资料.....	280
13.7 误用 t 检验处理析因设计的定量资料	281
13.8 误用 t 检验分析具有一个重复测量的单因素设计资料.....	282
13.9 误用 t 检验处理具有一个重复测量的 3 因素设计资料.....	283

13.10 误用 t 检验分析具有两个重复测量的 3 因素设计资料	284
13.11 误用 t 检验处理具有 3 个重复测量的 4 因素设计资料	285
13.12 误用配对设计资料的 t 检验处理单因素 k ($k \geq 3$) 水平设计资料	286
13.13 误用配对设计资料的 t 检验处理具有一个重复测量的单因素设计资料	286
13.14 误用配对设计资料的 t 检验处理具有一个重复测量的两因素设计资料	287
13.15 误用配对设计资料的 t 检验处理具有两个重复测量的两因素设计资料	290
13.16 误用 t 检验分析任意两组平均值之间的差异	291
13.17 误用 t 检验和一般方差分析处理带有协变量的资料	292
13.18 误用单因素多水平设计资料的方差分析处理析因设计资料	293
13.19 误用单因素多水平设计资料的方差分析处理具有一个重复测量的两因素设计资料	294
13.20 正交设计应用中存在的问题	295
第 14 章 定性资料分析中存在的统计学错误	298
14.1 误判资料性质, 从而误用定量资料统计分析方法	298
14.2 误用 t 检验处理定性资料	299
14.3 误用 t 检验分析四格表资料	300
14.4 误用 t 检验处理双向无序列联表资料	301
14.5 误用 t 检验处理单向有序列联表资料	303
14.6 单向有序资料未进行统计分析直接得出错误的结论	304
14.7 因 χ^2 值计算错误从而导致结论错误	304
14.8 根据略大于临界值的 χ^2 值直接拒绝原假设	305
14.9 误用 χ^2 检验取代 fisher 的精确检验	306
14.10 对 $R \times C$ 表资料直接分割进行两两比较	307
14.11 误用 χ^2 检验分析单向有序的列联表资料	308
14.12 误用 χ^2 检验回答相关性问题的	309
14.13 未进行统计分析便直接得出两有序变量的相关性	310
14.14 误用 χ^2 检验分析双向有序且属性相同的列联表资料	311
14.15 误用一般 χ^2 检验分析三维列联表资料	312
14.16 误用一般 χ^2 检验分析结果变量为有序变量的高维列联表资料	313
14.17 误用 Ridit 分析处理非单向有序列联表资料	314
14.18 误用秩和检验实现两有序变量之间的相关分析	316
14.19 双向有序列联表资料应根据分析目的选择相应的分析方法	317
14.20 误用 McNemar χ^2 检验处理不含“金标准”的配对设计定性资料	319
第 15 章 相关回归分析中存在的统计学错误	321
15.1 误对 I 型回归资料作相关分析	321

15.2 混淆相关与回归的概念	321
15.3 散布图并不反映直线趋势, 仍作直线相关分析	322
15.4 过小的 r 值在解释直线相关性时所遇到的问题	323
15.5 相关系数经假设检验所得的结论与原文结论相反	323
15.6 误用相关系数的假设检验来解释两检测方法的符合程度	324
15.7 用直线相关分析来回答其所不能回答的问题	324
15.8 误用一般 χ^2 检验来说明线性趋势问题	327
第 16 章 多元资料分析中存在的统计学错误	329
16.1 用一元分析取代多元分析	329
16.2 用单因素分析取代多因素分析	330
16.3 用单因素分析处理较复杂的资料	331
16.4 对生存资料不恰当的分析	333
16.5 多元回归分析中的策略错误	335
第 17 章 结果解释方面存在的统计学错误	337
17.1 不进行统计, 妄下结论	337
17.2 未明确交待统计分析方法, 得出结论	338
17.3 未交待资料类型及统计方法, 仅根据 P 值下结论	339
17.4 对结果的解释与所采用的统计分析方法不对应	339
17.5 两样本率的比较, 检验效能低	340
17.6 结果的解释脱离实际	341
17.7 对 P 值的误解	341
17.8 观察性研究推论因果关系	342
参考文献	344
附录 1 统计信息	347
附录 2 统计用表	352

第1篇 医学统计学学习策略

第1章 医学科研与论文中存在的 统计学问题及对策

1.1 统计学与科技事业的关系

1.1.1 统计学与科学研究的关系

2001年,笔者细读了《科技日报》上的两篇文章,即“科技头奖三届空缺的忧思”和“浮躁:科技界流行瘟疫”,深深感到这两篇文章写得太好了!大有发表太晚之感!在整个社会都显现出某种浮躁的大背景下,敢于从科技界“开刀”,说明人民对于科技界寄予很高的希望。事实上,类似的警钟在医学科研领域内早已敲响。尤其是从我国老一辈医学统计学专家开始,一代又一代的统计学工作者在一直不停地呼唤:各级领导、业务主管部门和广大的医学科研工作者,要高度重视科研工作的科学性和严谨性!

科学研究是人类探索未知事物变化发展规律的重要活动,而这些规律的发现,要靠科学的统计研究设计为指导,长期的观察和大量的实验为前提,对实验数据进行认真的分析和处理为依据,从而作出科学的推断和结论。在这个过程中,通常都需要运用统计学作为科学研究的得力助手,它使科学研究工作者头脑更清醒、智力更超常、眼睛更明亮、手段更有力。例如:在第二次世界大战期间,为了检验军火的质量,人类创造出了“序贯试验设计”,这一设计方法于20世纪50年代还成功地运用于医学研究领域;在1942年,英国将“正交设计”成功地用于马铃薯的肥料试验,随后,正交设计在日本的工业发展和兴起中起到了巨大的促进作用;20世纪70年代末,我国数学家王元和方开泰教授提出了“均匀设计”,这一设计方法为我国的导弹设计提供了科学的试验方案,取得了很好的应用效果。完善的、良好的试验设计方法,可以节省人力、物力、财力和时间;可以有效地控制和估计试验误差的大小;可以同时考察尽可能多的试验因素,从而达到经济、高效、可靠的试验目的。科学研究工作除了在研究设计方面需要运用统计学外,在资料的收集、整理、描述和表达,资料的分析,结果的解释,乃至专业结论的得出,也都离不开正确地运用统计学知识。因为任何一个环节上的疏忽或失误,都有可能导致错误的结论。

1.1.2 统计学与科研管理的关系

科研管理究竟应该管哪些？究竟应该怎么管？科研管理工作者整天忙忙碌碌，但如何才能培养出高水平的人才，如何才能创造出高质量的、富有创新性的成果，……，这些关系到我国科技事业兴衰存亡的大事，却很少有人去考虑。“过分注重直接的、短期的经济效益”已成为我国科技发展的羁绊，这一命题是对我国目前科技现状的真实写照。一些业务主管部门，对基础设施、基础教育和基础研究持蔑视和轻视的态度，对那些能在短期内产生直接经济效益的部门和个人，给予了过份的表彰。科研管理部门只考虑如何将大量的科研经费分配到各个研究单位去，却很少研究分配的政策和手段是否科学合理，很少研究投入的科研经费是否最大限度地发挥了应有的作用。事实上，在我国的科技工作中，许多科研课题的经费投入产出严重失调，其投入远远大于产出。

为什么会出现“投入远远大于产出”的现象呢？问题的症结就在于我国的科研管理水平太低，缺乏很有效的机制和对策去监督和管理科研课题。一个缺乏完善统计研究设计的科研课题，一个在学术上问题成堆的科研项目，通常也能顺利地通过成果鉴定，并获得高于其本身应有等级的成果。我们应该清楚地认识到，要想做好科研管理工作，应当充分发挥统计学的作用，对于科研课题的科学性的评价、科研经费的合理开支、投入产出的科学估算等，都离不开正确地运用统计学知识，仅凭经验和发号施令是低劣的管理水平。

1.1.3 统计学与科研成果的关系

与申请科研经费和课题一样，在科研成果的评审中，也存在“跑关系”的不良现象。有些领导对某些课题的技术问题一无所知，却偏偏是课题的主要负责人，其目的显而易见，即所申请的课题易于获准、经费易于获得、将来成果鉴定易于通过。这充分说明：在我国的课题申请和成果评审等工作中，存在很多人为干预的因素，存在很多不科学、不合理的问题。事实上，在成果的评审中，同样离不开统计学知识。看一个科研课题是否称职，首先就要看它是否有完善的统计研究设计方案；其次，要看它是否自始至终严格按设计方案具体实施；第三，要看它在资料收集、数据处理和得出结论等各个环节上，是否符合统计学的要求。很难想象，那些仅有专业人员组成的评审委员会，如何能对科研课题的学术水平和科学性好坏作出科学的、准确的、公正的评价！

1.1.4 统计学与科技期刊的关系

科技期刊是科研工作水平高低、科学性强弱的一面镜子，它可以反映出科研的成就，也可以暴露出科研中存在的问题。我国科技期刊，尤其是医学期刊中，存在大量的误用和滥用统计学的现象，这些问题将是本书的基本素材，由于它们的存在，给我国的科技期刊蒙上了一层令人羞愧的面纱。期刊论文中反映出来的问题，从实验设计、数据收集、统计分析方法