

口腔医学

科研设计与统计分析

KOUQIANG YIXUE
KEYAN SHEJI YU TONGJI FENXI

主编/胡良平

 人民军医出版社
PEOPLE'S MILITARY MEDICAL PRESS

口腔医学科研设计与统计分析

KOUQIANG YIXUE KEYAN SHEJI YU TONGJI FENXI

主 编 胡良平

编 委 (以姓氏笔画为序)

刘惠刚 李长平 沈艳红

罗艳侠 周诗国 单 彬

胡良平 彭迎春 葛 毅

编 者 (以姓氏笔画为序)

尹丽莉 刘 术 李 云

李 静 张 音 张 勇

张 倩 陈 浩 郝丽梅



人民军医出版社
People's Military Medical Press

北 京

图书在版编目(CIP)数据

口腔医学科研设计与统计分析/胡良平主编. —北京:人民军医出版社,2007.8

ISBN 978-7-5091-1170-3

I. 口… II. 胡… III. ①口腔科学—科学研究②口腔科学—医学统计—统计分析
IV. R78

中国版本图书馆 CIP 数据核字(2007)第 123165 号

策划编辑:于 岚 文字编辑:顾 森 责任审读:张之生

出版人:齐学进

出版发行:人民军医出版社

经销:新华书店

通信地址:北京市 100036 信箱 188 分箱

邮编:100036

质量反馈电话:(010)51927270;(010)51927283

邮购电话:(010)51927252

策划编辑电话:(010)51927300-8119

网址:www.pmmp.com.cn

印刷:北京国马印刷厂 装订:京兰装订有限公司

开本:787mm×1092mm 1/16

印张:22.5 字数:540千字

版、印次:2007年8月第1版第1次印刷

印数:0001~3000

定价:65.00元

版权所有 侵权必究

购买本社图书,凡有缺、倒、脱页者,本社负责调换

内 容 提 要

本书针对“许多人学了多遍统计学仍不得要领,几乎一用就出错”的普遍现象,提出了彻底解决的对策,其精髓就是“统计学三型理论(简称‘三型理论’)”,即统计学问题基本上都可归结为“表现型”、“原型”和“标准型”,准确把握每个具体问题中的“三型”,将能科学合理地解决科研工作中与统计学有关的实际问题。可以说,统计学中的全部内容皆可运用“三型理论”来解说,本书将灵活运用此理论为指导思想,辨析口腔医学科研中经常涉及到的“实验设计、统计描述、定量与定性资料的统计分析、相关与回归分析、多元统计分析”等内容在应用中常出现的错误,并介绍与前述内容有关的正确概念和做法。

书中给出了取自口腔医学科研设计和统计分析的大量实例,运用“三型理论”辨析“实验设计、统计描述和统计分析”中出现的错误,在给出正确做法的同时,还给出了带有原始数据的各种实例,用 SAS 软件演示统计分析的全过程和部分手工计算过程。为实际工作者科学合理地运用统计学提供了科学有效的理论和方法。

本书中绝大多数实例和资料取自口腔医学科研论文,但统计学的理论和方法具有普遍实用性。因此,本书不仅适合于口腔医学研究领域内的科研工作者、临床医生和学生,也适合于广大科技工作者、各科临床医生、杂志编辑、审稿专家和高等院校的师生学习和使用。

序

统计学是科学研究设计、资料数据处理、分析和表达的重要工具。应用得好,事半功倍。应用有误,功亏一篑。本人不仅仅把统计科学看作一种简单计算统计的工具,还把它看作在似乎纷乱无章,繁杂的数据中可能寻找到规律,乃至真理的指路明灯。统计学是每一个科研工作者必须掌握的一门工具学科。

在丰硕统计学专著中,由于深奥的数理原理和复杂的计算,使一般医学专业人员自学时不得要领,甚至知难而退。《口腔医学科研设计与统计分析》一书,不仅仅有对统计学理论的提纲挈领,还对在科学研究中经常要用的“随机原则”、“对照原则”,各种样本含量,单因素和多因素设计, t 检验, χ^2 检验和各种回归分析的概念、意义以及如何选择应用,都有深入浅出的表达。使人容易读懂。本书可贵之处是在论述这些理论、原则、概念和方法的同时,又有实例的对照。与通常的论著不同的是,这些实例都不是正面对照,而是“反面”对照。也许“反面”对照的教训让人懂得的更深刻而难忘。我想作者的原意在此而非彼。

主编胡良平教授是数理统计学专家,他一直努力把统计学原理应用到医学科学研究的实践中去,尤其是口腔医学科学研究的实践。多年来积累了丰富的经验和收集了许多宝贵的资料。虽然他不是学医学专业的,更不是口腔医学专业的,但是从他特有的数理统计视觉严格而尖锐地审视和评估口腔医学科学研究设计和统计中的一些问题,将对我们今后的工作是有益的。《口腔医学科研设计与统计分析》一书是我们口腔医学工作者的良师益友!

中华口腔医学会名誉会长
北京大学口腔医学院名誉院长

张震康

2007年5月20日

序

大家都知道,正确的统计学分析处理是医学科研工作能够得出科学结论并使之具有可信、可参考价值的基础。没有经过合理的正确的统计学分析处理的数据是难以评价其科学价值的。甚至作为科研论文的发表都会遇到障碍。然而正如胡良平教授所说,许多人学了多遍统计学仍不得要领,几乎一用就错。这固然是许多人没有学好统计学,即使学过,似乎还学得不错,也常常不会灵活运用则是另外一个重要原因。我本人也是这样,一遇到问题就不得不求助老师。

以往的统计学教科书大都是针对医学研究而编写的。专门关注口腔医学研究问题的统计学理论和实践,从口腔医学研究的实验设计、统计学表达和描述以及统计分析方法合理选择的角度并结合口腔医学研究中经常出现的统计学错误而加以分析指导的文章则较少。专著则更为罕见。

在这本专著中胡良平教授为彻底解决“学了统计学仍不得要领,一用就错”的普遍问题提出了对策,那就是掌握统计学三型理论,准确把握每个具体问题中的三型,即“表现型、原型和标准型”,从而科学合理地解决科研工作中与统计学有关的实际问题。胡教授在这本专著中就是以统计学三型理论为指导,并结合大量口腔医学研究中的实例,系统介绍在口腔医学研究中统计学应用的正确概念与做法。

我曾经多次聆听过胡良平教授关于医学研究中统计学问题研究的报告,同时他曾经担任《中华口腔医学杂志》的特邀编委,就口腔医学研究的统计学问题做过大量的深入研究,也就此做过许多报告。每次聆听他的报告都觉得受益匪浅。今天看到胡教授组织他的研究生花费5年时间完成了60万字的《口腔医学科设计与统计分析》的专著,更是感谢胡教授为中国口腔医学界的同仁们做了一件大好事。为正在从事口腔医学基础研究、临床研究的同事们提供了极具阅读和参考价值的好书。我相信有机会阅读这本专著的同事们一定会对提高自己的科研设计能力、统计表达及描述能力,统计分析能力有所帮助、有所补益。

我愿向读者们郑重推荐这本专著,希望我们的口腔医学研究工作水平不断提高,为我国口腔医学的发展做出更大贡献。

中华口腔医学会会长 王 兴

2007年6月11日于北京

序

随着科学技术的发展,医学科学包括口腔医学也在迅速发展,其中科学研究的进展尤为突出。

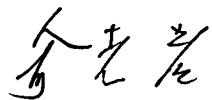
30多年前,在数量十分有限的口腔医学杂志上发表的论文基本上是回顾性临床经验总结,很少涉及科研设计。改革开放以后,迎来了科学的春天,培养了大批的研究生,一大批基础、应用基础以及前瞻性的研究应运而生。新的研究成果成批诞生,科研论文成倍增加,口腔医学杂志也迅速增多,口腔医学研究呈现出欣欣向荣的景象。

然而,由于相当一部分口腔医务工作者对于科研设计缺乏系统培训,对于统计学的知识缺乏深入了解,有的只是一知半解,不得要领;有的在理论上略知一二,但缺乏实际应用的经验。因而在论文中出现了不少缺陷,以至错误,明显影响了论文的质量,影响了科学研究结果的可靠性、真实性及可重复性,有的在一定程度上可能产生误导作用,这种现象必须引起口腔医学界同道的高度重视。

可喜的是,胡良平教授作为著名的生物医学统计学专家,多年来关注已发表的口腔医学科研论文,对于口腔医学科研设计和统计学分析进行了深入的研究,先后以会议专题讲座、举办学习班、在杂志上发表专家笔谈等多种形式向口腔医务工作者讲授口腔医学科研设计和统计分析,反应热烈。北京大学口腔医学院聘请胡良平教授为客座教授,胡教授经常给予理论讲授和实践指导,促进了我院科研工作的进展。我本人和课题组的同事也经常求教于他,获益匪浅。

胡良平教授组织研究生和相关统计学专家,经过为时5年的充分准备,完成了近60万字的《口腔医学科研设计与统计分析》一书。书中列举大量口腔医学科研设计和统计分析的实例,运用“统计学三型理论”辨析“实验设计、统计描述和统计分析”中出现的错误,给出正确的做法,为口腔医学工作者合理运用统计学提供了科学有效的理论和方法,这是一本难得的好书,是口腔医学百花园中的一朵绚丽鲜花,不仅对于口腔医学研究领域中的科研工作者、临床医生和学生起到直接的指导作用,而且可供杂志编辑和审稿专家借鉴和参考,无疑会对口腔医学事业的发展起到积极的推动作用。为此,我愿以满腔的热情将这本好书推荐给广大口腔医学界的同道。

北京大学口腔医学院院长



2007.5.6

前 言

数年前,我荣幸地结识了《中华口腔医学杂志》编辑部前任主任穆景燕编审和现任主任李季编审,通过他们的引荐,我又荣幸地结识了我国口腔界德高望重的老前辈、中华口腔医学会前任会长张震康教授和现任会长并兼任《中华口腔医学杂志》主编的王兴教授;北京大学口腔医学院的院领导俞光岩教授和马绪臣教授;《中华口腔医学杂志》名誉主编傅民魁教授和我国口腔界一大批精英,在向他们学习、与他们交流中,我学到了口腔科研方面的许多基本知识,逐渐引起了我对口腔科研论文的关注,并且使我对口腔医学科研中的科研设计与统计分析产生了浓厚兴趣。正是他们的不断鼓励和支持,使我有勇气和信心组织我的研究生和首都医科大学的几位同仁们共同撰写了《口腔医学科研设计与统计分析》这本书。

事实上,在5年前我们就开始着手收集、整理口腔科研论文中的科研设计和统计分析方面的资料。本书资料主要来源于我国口腔医学期刊。纵览国内外生物医学杂志,很难寻觅在科研设计和统计分析方面没有错误的论文,口腔医学杂志也不例外。我们撰写这本书的目的不是想诋毁任何杂志、杂志编辑部和原文作者,而是从科研工作的科学严谨角度来审视学术论文,将明显的、常识性的科研设计和统计分析错误展示出来,使广大实际工作者(即科研工作者、临床医生和杂志编辑等)充分意识到自己的研究工作中还有不少需要改进之处,我们大家都有责任和义务提高学术期刊的质量、提高科研的水平,从而提高广大实际工作者的科研素质。

笔者深深感谢张震康教授、王兴教授和俞光岩教授为本书写了序言,马绪臣教授在百忙之中打来电话表示关心和支持。

参与收集资料的人员主要是我院2002级研究生单彬、尹丽莉、刘术、李云、张勇、张音、张倩、陈浩、李静、郝丽梅等;参与编写的人员主要是我院生物医学统计学咨询中心的工作人员和研究生,他们是胡良平教授、周诗国讲师、李长平(在读博士)、葛毅(在读博士)、单彬(我中心培养的硕士研究生);首都医科大学的刘惠刚副教授、罗艳侠讲师、沈艳红讲师和彭迎春讲师。他们为本书的出版做出了卓越的贡献,在此,笔者向他们深表感谢!

尽管我们已竭尽全力,试图把这本书写好,但由于我们中的很多人是学数学和统计学的,在口腔专业方面是外行,主要依据统计学知识来审视问题,对某些具体问题的理解难免有偏倚,甚至是错误的,恳请口腔界的所有人士提出批评和建议,以便日后有机会再版时,能把错误降到最低限度,使本书真正成为口腔科研人员得心应手的工具书,为我国口腔事业的发展做出更大的贡献。

本书最终能够顺利问世,得益于口腔界前辈们的大力支持和扶植,得益于《中华口腔医学杂志》编辑部的热情关心和帮助,得益于《人民军医出版社》领导的高度重视和提携,得益于《人民军医出版社》彭倍勤主任和于岚编辑对出版事业的执著追求和忘我工作精神,得益于首都医科大学4位老师的鼎力相助。作为主编,笔者对所有为本书付出过辛勤劳动和给予精神鼓励的人们,致以最真诚的谢意!

胡良平

2007年7月15日

于北京军事医学科学院生物医学统计学咨询中心

目 录

第 1 章 统计学三型理论在口腔医学科研中的应用.....	(1)
1.1 统计学三型理论的定义及各型间的相互关系	(1)
1.1.1 什么是统计学三型理论	(1)
1.1.2 “三型”之间的相互关系	(2)
1.1.3 不易出错的“三型”和极易出错的“三型”	(2)
1.1.4 破解“三型”的技巧与策略	(2)
1.2 统计学三型理论产生的背景	(3)
1.3 统计学三型理论提出的依据	(4)
1.4 统计学三型理论适用的范围	(4)
1.5 统计学三型理论应用的样例	(5)
1.6 本章小结	(8)
第 2 章 统计表达和描述错误辨析与释疑.....	(9)
2.1 纵、横标目倒置.....	(9)
2.2 总纵标目设置错误.....	(14)
2.3 误用复式条图表达适合用线图表达的资料.....	(16)
2.4 误用复式条图表达构成比资料.....	(21)
2.5 用线图表达不宜用线图表达的资料.....	(21)
2.6 用普通线图表达适合用半对数线图表达的资料.....	(24)
2.7 计算相对数时分母太小.....	(24)
2.8 “率”和“比”混淆.....	(31)
2.9 误用表达呈正态分布资料的方法表达呈偏态分布的资料.....	(33)
2.10 用标准误描述数据的离散程度具有欺骗性	(42)
2.11 统计表中数据含义不清	(46)
2.12 在同一张表中表达多个不同的内容	(51)
2.13 统计图坐标轴上的刻度值不符合数学原则	(53)
2.14 随意列表,横标目重复.....	(60)
2.15 计算错误	(61)
2.16 数据错误	(62)
2.17 综合性错误	(63)
2.18 本章小结	(77)
第 3 章 实验设计及定量资料统计分析中的错误辨析与释疑	(78)

3.1 与随机原则有关的概念及实例	(78)
3.1.1 随机的概念与作用	(78)
3.1.2 违背随机原则的实例	(78)
3.2 与对照原则有关的概念及实例	(79)
3.2.1 对照的概念与作用	(79)
3.2.2 对照不全的概念	(79)
3.2.3 对照不全的实例	(80)
3.2.4 对照组的设立与实验设计目的不符的概念及后果	(86)
3.2.5 对照组的设立与实验设计目的不符的实例	(86)
3.2.6 重复设置对照组的概念	(90)
3.2.7 重复设置对照组的实例	(90)
3.2.8 盲目设置对照组的概念	(99)
3.2.9 盲目设置对照组的实例	(99)
3.2.10 实验分组不合理的概念	(102)
3.2.11 实验分组不合理的实例	(102)
3.3 与重复原则有关的概念与实例	(107)
3.3.1 重复原则的概念与作用	(107)
3.3.2 样本含量过小的实例	(107)
3.4 与均衡原则有关的概念及实例	(108)
3.4.1 均衡原则的概念	(108)
3.4.2 违背均衡原则的实例	(109)
3.5 与设计类型有关的概念及实例	(112)
3.5.1 设计类型的概念	(112)
3.5.2 误用单因素设计取代多因素设计	(112)
3.5.3 误将某种多因素设计判断为单因素设计而误用统计分析方法	(113)
3.6 实验设计中的其他错误实例	(114)
3.6.1 缺乏完善的设计方案	(114)
3.6.2 不合理的实验设计	(116)
3.6.3 脱离实际的实验设计方案	(117)
3.6.4 适合采用裂区设计的实验研究问题	(119)
3.6.5 实验设计中统计学错误的综合分析	(119)
3.7 三类常用实验设计类型的合理选用	(125)
3.7.1 合理选用单因素设计类型	(125)
3.7.2 合理选用多因素析因设计类型	(125)
3.7.3 合理选用多因素重复测量设计类型	(126)
3.8 本章小结	(126)
第4章 定性资料统计分析错误辨析与释疑	(127)
4.1 误将“例次”当作“例数”	(127)

4.2	定性资料的表达与分析显得很杂乱	(129)
4.3	列表格式和分析方法都不当	(130)
4.4	列表格式错误导致设计类型和资料类型发生变异	(131)
4.5	没有把调查目的如实地陈述清楚	(133)
4.6	误用定量资料统计分析方法处理定性资料	(134)
4.7	所交代的统计分析方法与所用的检验统计量的符号不吻合	(134)
4.8	忽视 χ^2 检验的前提条件且计算有误	(136)
4.9	误用 t 检验分析四格表资料	(136)
4.10	用一般 χ^2 检验取代校正 χ^2 检验或 Fisher 精确检验	(138)
4.11	用校正 χ^2 检验取代 Fisher 精确检验	(139)
4.12	用配对 χ^2 检验分析多个横断面研究设计的 2×2 列联表资料	(140)
4.13	误用一般 χ^2 检验取代 McNemar χ^2 检验	(141)
4.14	误用 χ^2 检验、 t 检验处理单向有序列联表资料	(142)
4.15	用多次两两比较随意割裂 $R \times C$ 表资料	(145)
4.16	用不恰当的列表格式处置一致性评价问题	(146)
4.17	误用 χ^2 检验处理三维列联表资料	(147)
4.18	误用 Ridit 分析处理三维列联表资料	(149)
4.19	多重 logistic 回归分析未完成	(150)
4.20	本章小结	(151)
第 5 章	定性资料的统计分析	(152)
5.1	横断面研究设计的 2×2 表资料及其统计分析	(152)
5.1.1	何为横断面研究设计的 2×2 表资料	(152)
5.1.2	实例分析	(153)
5.1.3	SAS 程序与结果解释	(153)
5.2	队列研究设计的 2×2 表资料及其统计分析	(154)
5.2.1	何为队列研究设计的 2×2 表资料	(154)
5.2.2	实例分析	(155)
5.2.3	SAS 程序与结果解释	(156)
5.2.4	值得注意的问题	(157)
5.3	病例对照研究设计的 2×2 表资料及其统计分析	(157)
5.3.1	何为病例对照研究设计的 2×2 表资料	(157)
5.3.2	实例分析	(158)
5.3.3	SAS 程序与结果解释	(158)
5.4	配对研究设计的 2×2 表资料及其统计分析	(160)
5.4.1	何为配对研究设计的 2×2 表资料	(160)
5.4.2	实例分析	(160)
5.4.3	SAS 程序与结果解释	(161)
5.5	双向无序的 $R \times C$ 表资料及其统计分析	(163)

5.5.1	何为双向无序的 $R \times C$ 表资料	(163)
5.5.2	实例分析	(164)
5.5.3	SAS 程序与结果解释	(164)
5.6	结果变量为有序变量的单向有序 $R \times C$ 表资料及其统计分析	(165)
5.6.1	何为结果变量为有序变量的单向有序 $R \times C$ 表资料	(165)
5.6.2	实例分析	(165)
5.6.3	SAS 程序与结果解释	(166)
5.7	双向有序且属性不同的 $R \times C$ 表资料及其统计分析	(168)
5.7.1	何为双向有序且属性不同的 $R \times C$ 表资料	(168)
5.7.2	实例分析	(169)
5.7.3	SAS 程序与结果解释	(169)
5.8	双向有序且属性相同的 $R \times C$ 表资料及其统计分析	(173)
5.8.1	何为双向有序且属性相同的 $R \times C$ 表资料	(173)
5.8.2	实例分析	(173)
5.8.3	几点说明	(177)
5.9	$2 \times K$ 列联表资料的统计分析	(177)
5.9.1	何为 $2 \times K$ 列联表资料	(177)
5.9.2	实例分析	(177)
5.10	$K \times 2$ 列联表资料的统计分析	(178)
5.10.1	何为 $K \times 2$ 列联表资料	(178)
5.10.2	实例分析	(179)
5.11	高维列联表资料的统计分析	(181)
5.11.1	何为高维列联表资料	(181)
5.11.2	实例分析	(181)
5.12	本章小结	(190)
第 6 章	定量资料的统计分析	(191)
6.1	单组设计定量资料的 t 检验与符号秩和检验	(192)
6.1.1	单组设计定量资料的 t 检验	(192)
6.1.1.1	单组设计的特点及资料应满足的前提条件	(192)
6.1.1.2	用 t 检验实现单组设计定量资料的统计分析	(192)
6.1.2	单组设计定量资料的符号秩和检验	(196)
6.1.2.1	采用符号秩和检验时资料应满足的前提条件	(196)
6.1.2.2	用符号秩和检验实现单组设计定量资料的统计分析	(196)
6.2	配对设计定量资料的 t 检验与符号秩和检验	(199)
6.2.1	配对设计定量资料的 t 检验	(199)
6.2.1.1	配对设计的特点及资料应满足的前提条件	(199)
6.2.1.2	用 t 检验实现配对设计定量资料统计分析	(200)
6.2.2	配对设计定量资料的符号秩和检验	(202)

6.2.2.1	采用符号秩和检验时资料应满足的前提条件	(202)
6.2.2.2	用符号秩和检验实现配对设计定量资料的统计分析	(202)
6.3	成组设计定量资料的 t 检验与秩和检验	(206)
6.3.1	成组设计定量资料的 t 检验	(206)
6.3.1.1	成组设计的特点及资料应满足的前提条件	(206)
6.3.1.2	用 t 检验实现成组设计定量资料的统计分析	(206)
6.3.2	成组设计定量资料的秩和检验	(210)
6.3.2.1	采用秩和检验时资料应满足的前提条件	(210)
6.3.2.2	用秩和检验实现成组设计定量资料的统计分析	(210)
6.4	单因素 $k(k \geq 3)$ 水平设计定量资料的方差分析与秩和检验	(215)
6.4.1	单因素 $k(k \geq 3)$ 水平设计定量资料的方差分析	(215)
6.4.1.1	对定量资料进行方差分析的前提条件	(215)
6.4.1.2	方差分析的基本思想	(215)
6.4.1.3	单因素 $k(k \geq 3)$ 水平设计定量资料的方差分析	(216)
6.4.2	单因素 $k(k \geq 3)$ 水平设计定量资料的秩和检验	(220)
6.4.2.1	采用秩和检验时资料应满足的前提条件	(220)
6.4.2.2	用秩和检验实现单因素 $k(k \geq 3)$ 水平设计定量资料的统计分析	(221)
6.5	析因设计定量资料的方差分析	(227)
6.5.1	析因设计的概念及特点	(227)
6.5.2	析因设计定量资料的方差分析实例	(228)
6.6	重复测量设计定量资料的方差分析	(231)
6.6.1	重复测量设计的概念及特点	(231)
6.6.2	重复测量设计定量资料的方差分析实例	(231)
6.6.2.1	实例、统计分析及结果解释	(232)
6.6.2.2	需要特别说明的问题	(240)
6.7	本章小结	(241)
第 7 章	相关与回归分析错误辨析与释疑	(243)
7.1	将接受不同处理或条件不同的几组受试对象放在一起进行相关分析	(243)
7.2	滥用直线相关分析得出没有实际意义的结论	(245)
7.3	散布图并不反映直线趋势,仍做直线相关分析	(245)
7.4	误用 χ^2 检验来说明相关回归问题	(246)
7.5	误用直线相关分析代替 Spearman 等级相关分析	(248)
7.6	误用 χ^2 检验回答相关性问题的、误用简单相关分析取代多元分析	(250)
7.7	多重线性回归变量筛选策略错误及误用简单相关分析取代复相关分析	(251)
7.8	多重 logistic 回归分析错误辨析与释疑	(253)
7.8.1	多重 logistic 回归分析中自变量的赋值错误	(253)
7.8.2	多重 logistic 回归模型中包含无统计学意义的自变量	(254)
7.8.3	多重 logistic 回归分析变量筛选策略错误	(256)

7.9	COX 模型回归分析错误辨析与释疑	(258)
7.9.1	COX 模型回归分析中自变量的赋值错误	(258)
7.9.2	COX 模型回归分析中变量筛选策略错误	(259)
7.10	多重回归分析方法描述不明确	(260)
7.11	多重回归分析中变量筛选方法的描述错误	(260)
7.12	相关或回归分析问题综合性错误的辨析与释疑	(261)
7.13	本章小结	(268)
第 8 章	简单相关与回归分析	(270)
8.1	简单线性相关与回归分析概述	(270)
8.1.1	简单线性相关与回归分析的概念	(270)
8.1.2	直线相关与回归分析的正确步骤	(270)
8.2	简单线性相关分析的计算	(272)
8.2.1	定量资料的 Pearson 直线相关分析	(272)
8.2.2	定性资料的 Spearman 秩相关分析	(273)
8.3	简单线性回归分析的计算	(274)
8.3.1	截距 a 和斜率 b 的计算	(274)
8.3.2	截距 a 和斜率 b 的假设检验	(274)
8.3.3	总体截距 α 和斜率 β 的置信区间	(275)
8.3.4	简单线性回归分析中其他有关的区间估计问题	(275)
8.3.5	简单线性回归分析的作用	(276)
8.4	直线相关与回归分析的异同点	(277)
8.4.1	不同点	(277)
8.4.2	相同点	(277)
8.4.3	内在联系	(277)
8.5	简单线性相关与回归分析的实践与应用	(277)
8.6	本章小结	(283)
第 9 章	多重线性回归分析和多重 logistic 回归分析	(285)
9.1	多重线性回归分析	(285)
9.1.1	多重线性回归分析的应用场合	(285)
9.1.2	多重线性回归分析的概念和任务	(285)
9.1.3	多重回归模型中变量筛选的方法	(286)
9.1.4	回归诊断方法	(288)
9.1.5	多重回归模型优劣的评价标准	(289)
9.1.6	多重回归模型中各自变量作用大小的评价	(290)
9.1.7	多重线性回归分析应用实例	(290)
9.2	多重 logistic 回归分析	(296)
9.2.1	多重 logistic 回归分析的应用场合	(296)

9.2.2	多重 logistic 回归分析的概念、分类及任务	(296)
9.2.3	logistic 回归方程的参数估计及假设检验	(296)
9.2.4	logistic 回归方程变量筛选的方法	(297)
9.3	多重回归模型的合理选用与正确解释	(297)
9.4	多重回归分析的应用条件	(297)
9.4.1	多重线性回归分析的应用条件	(297)
9.4.2	多重 logistic 回归分析的应用条件	(297)
9.5	变量的数量化	(297)
9.6	多重 logistic 回归分析的实践与应用	(298)
9.7	本章小结	(302)
第 10 章	常见多因素实验设计类型辨析	(304)
10.1	生物医学科研中三种最常见的多因素实验设计类型概述	(304)
10.2	多因素非平衡的组合实验	(305)
10.3	多因素重复测量设计	(309)
10.3.1	对称部位充当“重复测量因素”	(309)
10.3.2	处理前后充当“重复测量因素”	(310)
10.4	多因素析因设计与重复测量设计交织在一起	(310)
10.5	本章小结	(319)
参考文献		(320)
附录 A	胡良平专著简介	(323)
附录 B	常用统计用表	(325)

第1章 统计学三型理论在口腔医学科研中的应用

很多人都知道统计学很难学,尤其是使用起来困难更大。“许多人学了多遍统计学仍不得要领,几乎一用就错”已成为不容置疑的事实。但为什么会这样,有什么好的解决办法,却很少有人去思考,更无人提出有效的解决方案。这正是导致全球生物医学科研设计和统计分析质量令人担忧的客观现实!口腔医学科研属于临床医学科研的一个分支,其科研设计与统计分析中也存在不少问题,应当引起有关人士的高度关注。本章将分析导致普遍误用统计学的根源,并创造性地提出可有效解决前述问题的办法,即“统计学三型理论”。在此基础上,结合口腔医学科研实际,通过实例展示如何正确运用统计学理论解决与实验设计、统计描述和统计分析方法合理选择有关的实际问题。

1.1 统计学三型理论的定义及各型间的相互关系

1.1.1 什么是统计学三型理论

统计学三型理论(简称“三型理论”)就是把科研工作中与统计学有关的问题归结为“表现型”、“原型”和“标准型”,从而有利于深刻揭示问题的本质,科学合理地运用统计学理论和方法解决各种实际问题的一种新理论。

笔者通过阐述“三型理论”的客观存在性及各型之间的相互关系,提出破解三型之谜的技巧与策略,为人们编写高质量的统计学教材、巧妙地讲授统计学、轻松地学习统计学和正确地应用统计学提供一种崭新的理念、理论和方法。统计学三型理论可以使实际工作者在实验设计类型的识别和统计分析方法的合理选用上起到举一反三、触类旁通的效果。

具体地讲,什么叫“表现型”呢?“表现型”就是实际工作者将一个与统计学有关的专业问题以自己最习惯的形式呈现出来的一种模式或结构,这种模式或结构常以一种假象出现,对合理选用统计分析方法处理资料常会起误导作用。比如说,某研究者同时用A、B两种药做某实验,每种药又考虑大小两种剂量,一共可以形成4个实验组,研究者习惯上就认为“组别”是此实验中的“实验因素”,在其下标出“第1组、第2组、第3组、第4组”,给出各组某些定量观测指标观测值的平均值和标准差,并误认为此实验设计是单因素四水平设计,接着进行6次 t 检验或做1次单因素四水平设计定量资料的方差分析和 q 检验,显然,这都是错误的!这是由于研究者被“表现型”的假象迷惑所致。

什么是“原型”呢?“原型”就是能全面正确反映实际工作者研究目的的一种模式或结构,这种模式或结构通常把问题的本质呈现出来了。例如,在前例中,在“药物分组”之下表示出“A和B药均用小剂量、A药小剂量B药大剂量、A药大剂量B药小剂量、A和B药均用大剂

量”，这样把各药物组的含义明确表达出来，做统计分析时就会很慎重，一般不会盲目去进行两两比较。

什么是“标准型”呢？“标准型”就是统计学教科书上对各类问题习惯上采用的一种表达模式或结构，这种模式或结构常以“不言自明”的方式把问题的本质呈现出来。例如，在前例中，不用“组别”或“药物分组”等字样，而用“A药剂量”与“B药剂量”这样两个词，在统计学上称他们为两个实验因素，每个实验因素都有大剂量和小剂量两个水平，与此实验对应的实验设计自然就是“两因素设计”了，更确切地说，应叫做两因素析因设计或 2×2 析因设计（注意：在两因素实验中，根据某些假设，还有其他的实验设计类型，如两因素系统分组设计、两因素分割设计等，因篇幅所限，此处不详述）。

1.1.2 “三型”之间的相互关系

通过分析和总结医学科研工作中出现的大量实际问题，不难发现：有些问题的“表现型”就是问题的“原型”；有些问题的“表现型”需要通过结构变形使其转变成“原型”；还有些问题的“表现型”需要通过拆分使其转变成“原型”。

“原型”与“标准型”之间存在什么样的关系呢？有些问题的“原型”就是问题的“标准型”；有些问题的“原型”需要分解成多个“标准型”；还有些问题的“原型”根本不存在与之对应的“标准型”，这种情形通常发生在实验设计不规范、甚至有严重错误的场合中。此时，最好是推翻原先的设计，重新设计，重做实验。

1.1.3 不易出错的“三型”和极易出错的“三型”

若与一个实际问题对应的“三型”完全相同，只要这个问题本身不很复杂，而且实际工作者已具备处理此类问题所需要的知识，则处理这样的问题时通常是不易出错的。

若与一个实际问题对应的“三型”属于上述最复杂的情形，即“表现型”需要经过变形或拆分才能转变成“原型”，若“原型”又根本不存在与之对应的“标准型”，则实际工作者处理这样的问题时是极易出错的。

1.1.4 破解“三型”的技巧与策略

全面学习和掌握各类问题的“标准型”，学会透过“表现型”的表象看清其“原型”的本质，借助专业知识和统计学知识将“原型”所对应的“标准型”（如果存在的话）揭示出来，从而使问题迎刃而解；若“原型”所对应的“标准型”根本不存在，则对拟解决的问题做出应有的裁决（很可能实验设计有严重错误或数据无法得到正确的处理）。

怎样才能学好统计学三型理论呢？首先需要对统计学中各种问题所对应的“标准型”有全面系统的了解和掌握，力求做到“胸有成竹、遇事不慌”；其次，要冷静地看待实际工作者（包括学习者本人）习惯采用的“表现型”，不要被表象所迷惑；第三，要紧密结合专业知识和统计学知识将实际问题的本质搞清楚，使问题的“原型”能更真实地显露出来。对于拟用统计学解决的每一个具体问题，先从问题的“表现型”入手，再设法弄清问题的“原型”，进而将“原型”通过变形或拆分使其正确地转变成统计学上的“标准型”，以使用最合适的统计学方法予以处理。