

XINXUEGUANBING
KEYAN SHEJI YU TONGJI FENXI

主编 胡良平

心血管病 科研设计与统计分析

很多与统计学有关的实际问题，均以“表现型”的面貌呈现在人们的面前，表现型常常带有假象，直接依据表现型去盲目套用传统的统计学教科书上的“标准型”，十有八九会出错，因此，要想正确运用统计学，必须弄清反映“表现型”本质的“原型”，将“原型”正确转变成“标准型”后，就很少会出错。这样一种由胡良平创立的可有效解决问题的新理论，被称为“统计学三型理论”。此理论可使统计学思想付诸实施。

统计学三型理论



人民军医出版社

PEOPLE'S MILITARY MEDICAL PRESS

心血管病科研设计与统计分析

XINXUEGUANBING KEYAN SHEJI YU TONGJI FENXI

主 编 胡良平

副主编 柳伟伟 高 辉

编 委 (以姓氏笔画为序)

周诗国 胡良平 柳伟伟

高 辉 郭 晋 程瑞专



人民军医出版社

PEOPLE'S MILITARY MEDICAL PRESS

北 京

图书在版编目(CIP)数据

心血管病科研设计与统计分析/胡良平主编. —北京:人民军医出版社,2010.6
ISBN 978-7-5091-3852-6

I. ①心… II. ①胡… III. ①心脏血管疾病—研究②心脏血管疾病—统计分析 IV. ①R54

中国版本图书馆 CIP 数据核字(2010)第 094885 号

策划编辑:于 岚 文字编辑:仲静洁 吴 倩 责任审读:黄栩兵

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)51927290;(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:三河市祥达印装厂 装订:京兰装订有限公司

开本:787mm×1092mm 1/16

印张:20 字数:475 千字

版、印次:2010 年 6 月第 1 版第 1 次印刷

印数:0001~3500

定价:60.00 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

编者根据我国数十种学术期刊发表的心血管疾病科研论文的科研设计情况及统计问题：首先从正面讲述统计学的主要内容，包括统计表达与描述、实验设计、定量与定性资料统计分析、简单相关回归分析和多重回归分析；紧接着围绕这些内容，针对人们误用统计学的实际案例，对差错进行辨析与释疑。大部分内容与心血管疾病的基础研究、疾病诊治和预防有关，但其解决问题的思维方法和统计学方法适合于任何生物医学基础研究和临床研究。在正面讲述统计学中，凡涉及较复杂的统计计算问题，都给出用国际上著名统计分析系统 SAS 实现计算的方法（即提供了 SAS 引导程序）和程序结果及解释。本书适合于从事生物医学基础研究和临床研究的科研工作者、医师、本科生、研究生和杂志编辑学习与参考，也适合于一切需要学习和运用 SAS 软件解决各种常用统计分析问题，以及希望了解和掌握实验设计和统计表达与描述知识的人们。

前 言

随着人民生活水平的日益提高,我国人民面临着一些新出现的问题,其中,心血管疾病的发病率呈逐年上升趋势。心血管疾病越来越严重地威胁着人类的健康,广大医学工作者正致力于寻找导致心血管疾病发生和发展的原因,并努力寻找对策。无论是寻找原因,还是对策,都离不开艰苦卓绝的科学研究,这是因为导致疾病发生的原因很多,有效的治疗方案不是轻而易举就可以获得的。人们需要在众多的致病原因中找出起决定作用的原因,要从多种可能的治疗方案中优选出最佳的治疗方案,这些都属于多因素、多指标的科学研究问题。

如何才能科学严谨、经济高效地进行和完成多因素、多指标的科学研究问题呢?第一,研究者需要掌握丰富的医学、生物学、遗传学等相关学科的基本常识和专业常识;第二,需要有大量的临床工作经验,并善于总结古今中外的知识、经验和教训;第三,需要具备较丰富的统计学知识。在统计学知识方面,最主要的是科研设计知识和技能,其核心内容是在充分而全面地利用基本常识和专业常识的基础上,合理地考虑和安排三要素(受试对象、影响因素和观测指标)、四原则(随机、对照、重复和均衡)及实验设计类型;其次,正确地表达和描述实验资料,对准确揭示事物内在的规律性也是很有必要的;再次,就是合理选用统计分析方法处理定量和定性资料,揭示变量之间相互和依赖关系。

仅从正面入手学习统计学往往收效甚微,一到用时就很容易出错。为此,本书在上篇从正面讲授常用统计学知识(共5章,即统计表达与描述、实验设计、定量资料统计分析、定性资料统计分析、简单相关回归分析及多重回归分析)的基础上,又用约全书一半的篇幅介绍心血管病科研中人们在统计学方面所犯的形形色色的错误(后5章的名称与上篇中5章的基本相同,但内容是针对人们误用统计学的实例进行辨析与释疑),使学习者“吃一堑,长一智”,从他人“错误的经历”中汲取经验教训。

第1章和第6章属于统计学基础知识,是从正反两个方面讲授统计表达与描述方面的内容,由程瑞专完成初稿;第2章和第7章属于实验设计方面的知识,是从正反两个方面介绍实验设计方面的内容,由周诗国完成初稿;第3章和第8章属于定量资料统计分析方面的知识,是从正反两个方面介绍处理定量资料方面的内容,由高辉完成初稿;第4章和第9章属于定性资料统计分析方面的知识,是从正反两个方面介绍处理定性资料方面的内容,由郭晋完成初稿;第5章和第10章属于简单相关回归与多重回归分析方面的知识,也是从正反两个方面介绍如何研究变量之间相互关系和依赖关系的内容,由柳伟伟完成初稿。

在本书即将出版之际,笔者要衷心感谢撰写初稿的5位青年教师和研究生,特别要感谢柳伟伟讲师,他协助笔者为本书初稿做了大量审阅工作,以及博士生高辉,他为本书高质量地完成

了约 40% 的书稿,还要感谢本中心在读硕士研究生毛玮、陶丽新、王琪、鲍晓蕾、贾元杰,他们为本书作了认真校对。最后,要感谢所有为本书审校和出版付出过辛勤劳动的人们。

由于笔者水平有限,不妥或错误之处,恳请广大读者不吝赐教,以便再版时修正。

主编 胡良平

北京军事医学科学院生物医学统计学咨询中心

2010 年 1 月 25 日

目 录

上篇 统计学内容概要

第 1 章 统计表达与描述	(3)
1.1 资料类型	(3)
1.1.1 现代划分方法	(3)
1.1.2 划分方法的本质	(4)
1.1.3 资料类型转换	(4)
1.2 定量资料的统计描述	(5)
1.2.1 对称分布资料	(6)
1.2.2 非对称分布资料	(8)
1.2.3 其他常用描述性统计指标	(10)
1.3 定性资料的统计描述	(13)
1.3.1 二分类资料	(13)
1.3.2 多分类资料	(13)
1.3.3 相对数	(14)
1.4 正态分布及其应用	(17)
1.4.1 概念	(17)
1.4.2 性质	(18)
1.4.3 应用	(20)
1.5 统计表	(23)
1.5.1 构成及要求	(23)
1.5.2 种类	(23)
1.6 统计图	(24)
1.6.1 构成及要求	(24)
1.6.2 常用种类及应用	(24)
1.7 本章小结	(29)
第 2 章 实验设计	(30)
2.1 科研设计概述	(30)
2.1.1 概念与目的	(30)
2.1.2 内容	(30)
2.2 统计研究设计概述	(30)
2.2.1 要点及注意事项	(30)
2.2.2 设计方案是否完善的评价标准	(31)

2.3 统计研究设计类型概况	(31)
2.3.1 观察性研究设计	(32)
2.3.2 实验性研究设计	(32)
2.4 统计研究设计中的概念	(37)
2.4.1 因素与水平	(37)
2.4.2 组别与处理	(38)
2.4.3 指标、项目与效应	(39)
2.4.4 变量、自变量与因变量	(39)
2.4.5 交互作用、协同作用与拮抗作用	(40)
2.4.6 平衡设计与非平衡设计	(41)
2.5 实验设计要点概述	(41)
2.5.1 意义	(42)
2.5.2 原理	(42)
2.5.3 三要素	(42)
2.5.4 四原则	(43)
2.5.5 设计类型	(44)
2.6 树立正确的科研工作指导思想	(45)
2.7 本章小结	(48)
第3章 定量资料统计分析	(50)
3.1 单组设计	(50)
3.1.1 设计类型	(50)
3.1.2 统计分析方法	(51)
3.1.3 实例分析	(53)
3.2 配对设计	(55)
3.2.1 设计类型	(55)
3.2.2 统计分析方法	(55)
3.2.3 实例分析	(55)
3.3 成组设计	(57)
3.3.1 设计类型	(57)
3.3.2 统计分析方法	(58)
3.3.3 实例分析	(59)
3.4 单因素 $k(k \geq 3)$ 水平设计	(64)
3.4.1 设计类型	(64)
3.4.2 统计分析方法	(64)
3.4.3 实例分析	(65)
3.5 单因素设计的其他内容	(71)
3.5.1 正态性检验	(71)
3.5.2 方差齐性检验	(73)
3.5.3 单因素 k 水平设计定量资料组间均数差别的多重比较	(73)
3.6 随机区组设计	(74)

3.6.1 设计类型	(74)
3.6.2 统计分析方法	(74)
3.6.3 实例分析	(75)
3.7 析因设计	(79)
3.7.1 设计类型	(79)
3.7.2 统计分析方法	(80)
3.7.3 实例分析	(81)
3.8 嵌套设计	(84)
3.8.1 设计类型	(84)
3.8.2 统计分析方法	(85)
3.8.3 实例分析	(86)
3.9 重复测量设计	(93)
3.9.1 设计类型	(93)
3.9.2 统计分析方法	(93)
3.9.3 实例分析	(94)
3.10 本章小结	(110)
第4章 定性资料统计分析	(112)
4.1 横断面研究设计的 2×2 表资料	(112)
4.1.1 特点及分析方法	(112)
4.1.2 实例分析	(113)
4.1.3 SAS 程序与结果解释	(113)
4.2 队列研究设计的 2×2 表资料	(115)
4.2.1 特点及分析方法	(115)
4.2.2 实例分析	(115)
4.2.3 SAS 程序与结果解释	(116)
4.2.4 值得注意的问题	(117)
4.3 病例-对照研究设计的 2×2 表资料	(117)
4.3.1 特点及分析方法	(117)
4.3.2 实例分析	(118)
4.3.3 SAS 程序与结果解释	(118)
4.4 配对研究设计的 2×2 表资料	(119)
4.4.1 特点及分析方法	(119)
4.4.2 实例分析	(120)
4.4.3 SAS 程序与结果解释	(120)
4.5 双向无序的 $R \times C$ 表资料	(122)
4.5.1 特点及分析方法	(122)
4.5.2 实例分析	(122)
4.5.3 SAS 程序与结果解释	(123)
4.6 结果变量为有序变量的单向有序 $R \times C$ 表资料	(123)
4.6.1 特点及分析方法	(123)

4.6.2	实例分析	(124)
4.6.3	SAS 程序与结果解释	(124)
4.7	原因变量为有序变量结果变量为二值变量的 $R \times C$ 表资料	(125)
4.7.1	特点及分析方法	(125)
4.7.2	实例分析	(126)
4.7.3	SAS 程序与结果解释	(126)
4.8	双向有序且属性不同的 $R \times C$ 表资料	(127)
4.8.1	特点及分析方法	(127)
4.8.2	实例分析	(128)
4.8.3	SAS 程序与结果解释	(128)
4.9	双向有序且属性相同的 $R \times C$ 表资料	(130)
4.9.1	特点及分析方法	(130)
4.9.2	实例分析	(131)
4.9.3	SAS 程序与结果解释	(131)
4.9.4	需要注意的几点问题	(133)
4.10	结果变量为二值变量的高维列联表资料	(133)
4.10.1	特点及分析方法	(133)
4.10.2	实例分析	(133)
4.10.3	SAS 程序与结果解释	(134)
4.11	结果变量为多值名义变量的高维列联表资料	(137)
4.11.1	特点及分析方法	(137)
4.11.2	实例分析	(137)
4.11.3	SAS 程序与结果解释	(137)
4.12	结果变量为多值有序变量的高维列联表资料	(140)
4.12.1	特点及分析方法	(140)
4.12.2	实例分析	(140)
4.12.3	SAS 程序与结果解释	(140)
4.13	具有重复测量设计的高维列联表资料	(142)
4.13.1	特点及分析方法	(142)
4.13.2	实例分析	(142)
4.13.3	SAS 程序与结果解释	(143)
4.13.4	注意事项	(144)
4.14	本章小结	(144)
第 5 章	简单相关回归分析及多重回归分析	(145)
5.1	简单线性相关与回归	(145)
5.1.1	基本概念	(145)
5.1.2	简单线性相关分析	(145)
5.1.3	Spearman 秩相关分析	(146)
5.1.4	简单线性回归分析	(147)
5.1.5	简单线性相关和回归的联系与区别	(149)

5.1.6 注意事项	(150)
5.1.7 应用举例	(151)
5.2 多重线性回归	(154)
5.2.1 基本概念	(154)
5.2.2 回归方程的计算	(154)
5.2.3 回归方程和偏回归系数的假设检验	(154)
5.2.4 自变量的筛选方法	(155)
5.2.5 回归诊断方法	(157)
5.2.6 自变量作用大小的评价	(158)
5.2.7 应用举例	(158)
5.3 多重 Logistic 回归	(164)
5.3.1 基本概念	(164)
5.3.2 参数估计和假设检验	(165)
5.3.3 自变量的筛选方法	(165)
5.3.4 回归模型的拟合优度检验	(165)
5.3.5 应用举例	(166)
5.4 本章小结	(169)

下篇 统计学应用中常见错误辨析与释疑

第 6 章 统计表达与描述错误辨析与释疑	(173)
6.1 误用有序资料取代定量资料	(173)
6.2 平均指标与变异指标应用中存在的问题	(174)
6.2.1 误用正态分布法描述偏态分布资料	(174)
6.2.2 误用标准误作为变异指标描述定量资料	(175)
6.3 相对数应用中存在的问题	(176)
6.3.1 计算相对数分母过小	(176)
6.3.2 比和率混淆使用	(176)
6.4 统计表中存在的问题	(177)
6.4.1 主谓语颠倒	(177)
6.4.2 主次标目颠倒	(178)
6.4.3 绝对数与相对数交叉出现	(178)
6.4.4 计算相对数时分母过小	(179)
6.5 统计图中存在的问题	(180)
6.5.1 误用复式条图表达适合用线图表达的资料	(180)
6.5.2 误用普通线图表达适合绘制条图的资料	(181)
6.5.3 条图的直条间未留空隙	(183)
6.6 本章小结	(184)
第 7 章 实验设计错误辨析与释疑	(186)
7.1 与“三要素”有关的实验设计错误案例	(186)

7.1.1	受试对象的交代不具体,缺乏样本含量的概念	(186)
7.1.2	受试对象的选取违背同质性原则,不具有代表性	(187)
7.1.3	选取的部分受试对象与研究目的无关	(188)
7.2	与“四原则”有关的实验设计错误案例	(188)
7.2.1	实验设计违反随机原则	(188)
7.2.2	设置了多余的对照组	(190)
7.2.3	未设置对照组	(196)
7.2.4	实验设计违反均衡原则	(197)
7.2.5	实验设计对照不全	(201)
7.3	与“设计类型”有关的实验设计错误案例	(214)
7.3.1	误用局部设计代替整体设计	(214)
7.3.2	缺乏实验设计类型的概念	(218)
7.4	本章小结	(219)
第8章	定量资料统计分析方法应用中的错误辨析与释疑	(224)
8.1	误用 χ^2 检验处理定量资料	(224)
8.2	忽视参数检验应用的前提条件	(226)
8.3	误用配对设计定量资料 t 检验处理成组设计定量资料	(228)
8.4	误用 t 检验处理单因素 $k(k \geq 3)$ 水平设计定量资料	(229)
8.5	误用配对设计定量资料 t 检验处理具有一个重复测量的单因素设计定量资料	(230)
8.6	误用配对设计定量资料 t 检验处理具有一个重复测量的两因素设计定量资料	(232)
8.7	误用成组设计定量资料 t 检验处理具有一个重复测量的两因素设计定量资料	(233)
8.8	误用成组设计定量资料 U 检验处理具有一个重复测量的两因素设计定量资料	(235)
8.9	误用配对设计和成组设计定量资料 t 检验处理具有一个重复测量的两因素设计定量资料	(235)
8.10	误用成组设计定量资料 t 检验处理具有一个重复测量的三因素设计定量资料	(238)
8.11	误用配对设计定量资料 t 检验处理具有两个重复测量的两因素设计定量资料	(238)
8.12	误用成组设计定量资料 t 检验处理具有两个重复测量的两因素设计定量资料	(239)
8.13	误用 t 检验处理具有两个重复测量的三因素设计定量资料	(240)
8.14	误用 t 检验处理两因素析因设计或嵌套设计定量资料	(242)
8.15	误用 t 检验处理三因素析因设计或嵌套设计定量资料	(244)
8.16	误用 t 检验处理多因素非平衡组合实验定量资料	(245)
8.17	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析及 t 检验处理具有一个重复测量的两因素设计定量资料	(248)
8.18	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析处理具有一个重复测量的三因素设计定量资料	(249)
8.19	误用 SNK 检验处理具有一个重复测量的三因素设计定量资料	(250)

8.20	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析处理具有两个重复测量的两因素设计定量资料	(251)
8.21	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析处理两因素析因设计或嵌套设计定量资料	(252)
8.22	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析处理三因素析因设计或嵌套设计定量资料	(255)
8.23	误用单因素 $k(k \geq 3)$ 水平设计定量资料方差分析处理多因素非平衡组合实验定量资料	(256)
8.24	数据不符合数学逻辑	(259)
8.25	统计分析计算错误	(260)
8.26	本章小结	(261)
第 9 章	定性资料统计分析方法应用中的错误辨析与释疑	(262)
9.1	忽视一般 χ^2 检验的前提条件	(262)
9.2	误用 t 检验处理定性资料	(263)
9.3	误用一般 χ^2 检验代替秩和检验处理结果变量为有序变量的 $R \times C$ 表资料	(263)
9.4	原因变量是有序变量结果变量为二值变量的 $R \times C$ 表资料分析中存在的错误	(265)
9.5	误用一般 χ^2 检验得出相关性结论	(267)
9.6	未使用统计分析方法而得出结论	(269)
9.7	误用 Spearman 秩相关分析双向有序且属性相同的 $R \times C$ 表资料	(270)
9.8	误用一般 χ^2 检验分析结果变量为二值变量的高维列联表资料	(271)
9.9	误用一般 χ^2 检验分析结果变量为多值有序变量的高维列联表资料	(273)
9.10	误用一般 χ^2 检验分析结果变量为多值名义变量的高维列联表资料	(274)
9.11	误用一般 χ^2 检验分析具有重复测量设计的高维列联表资料	(275)
9.12	误用 Fisher 精确检验分析具有重复测量设计的高维列联表资料	(278)
9.13	本章小结	(279)
第 10 章	简单相关回归及多重回归分析方法应用中的错误辨析与释疑	(280)
10.1	简单线性相关与回归中的错误辨析与释疑	(280)
10.1.1	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	(280)
10.1.2	散点图并不反映直线趋势,仍进行直线相关分析	(281)
10.1.3	对定性变量或非正态变量进行直线相关分析	(282)
10.1.4	应用直线相关分析得出没有实际意义的结论	(284)
10.1.5	误用简单相关分析取代复相关分析或多重线性回归分析	(285)
10.1.6	使用 Logistic 回归得出直线相关系数	(286)
10.2	多重回归中的错误辨析与释疑	(286)
10.2.1	使用多重线性回归分析结果变量为定性变量的资料	(286)
10.2.2	未对多值名义的自变量产生哑变量	(289)
10.2.3	最终的多重 Logistic 回归方程中包含无统计学意义的自变量	(290)
10.2.4	将单变量分析的结果作为筛选自变量的依据	(291)
10.2.5	多重回归分析中只使用一种自变量筛选方法	(294)
10.2.6	回归系数与优势比的估计值相矛盾	(295)

10.2.7 使用二值变量的 Logistic 回归处理结果变量为多值名义变量的资料	(296)
10.2.8 Logistic 回归中优势比 95% 置信区间的上限为无穷大	(297)
10.3 本章小结	(298)
附录 胡良平统计学专著及配套软件简介	(300)

上 篇

统计学内容概要

