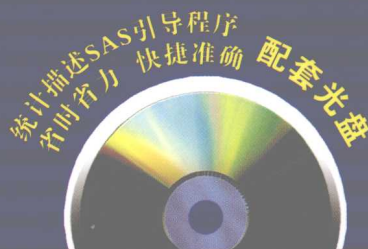


统计学三型理论

在统计表达与描述中的应用

TONGJIXUE SANXINGLILUN ZAI TONGJI BIAODA
YU MIAOSHU ZHONG DE YINGYONG

主编 胡良平



生物医学研究和论文写作离不开正确的统计表达与描述，离不开使用SAS软件。长期以来，使用SAS软件者需要花费大量的时间和精力，将每个SAS程序的大量语句、符号录入计算机，反复核对、调试正常后才能正确调用SAS软件包。

统计描述SAS引导程序，是我国统计学专家胡良平教授等多位统计、计算机专业工作者用近一年时间潜心钻研SAS软件包中有关概率分布及其图形绘制方法后编写而成的，为国内专门用于统计表达与描述的第一个SAS引导程序。

读者通过此光盘，可充分发挥计算机软件的功能和作用，用鼠标左键双击文件名，轻松实现SAS程序的调用；用新数据替换SAS引导程序中的老数据，单击发送键即可完成统计表达。该程序可反复使用，省时省力，快捷准确，使统计表达与描述方面的问题均能得到解决。本盘适于学习和使用统计学及撰写各类统计报告、文书的人员。

统计学三型理论

在统计表达与描述中的应用

TONGJIXUE SANXINGLILUN
ZAI TONGJI BIAODA YU MIAOSHU ZHONG DE YINGYONG

主 编 胡良平

编 者 (以姓氏笔画为序)

王诗鸿 刘惠刚 李 鹏 李子建

沈艳红 周诗国 胡良平 胡纯严

柳伟伟 高 辉 彭迎春 葛 毅

程瑞专

秘 书 李长平 高 辉 郭 晋

 人民军医出版社

PEOPLE'S MILITARY MEDICAL PRESS

北 京

图书在版编目 (CIP) 数据

统计学三型理论在统计表达与描述中的应用/胡良平主编.

—北京: 人民军医出版社, 2008.6

ISBN 978-7-5091-1871-9

I. 统… II. 胡… III. 统计学—基本知识 IV.C8

中国版本图书馆 CIP 数据核字 (2008) 第 081556 号

策划编辑: 于 岚 贝 丽 文字编辑: 赵晶辉 责任审读: 李 晨

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) 51927273

网址: www.pmmp.com.cn

印刷: 北京京海印刷厂 装订: 京兰装订有限公司

开本: 787mm×1092mm 1/16

印张: 24 字数: 553 千字

版、印次: 2008 年 6 月第 1 版第 1 次印刷

印数: 0001~3000

定价: 80.00 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

统计表达与描述是统计学的基础，无论多么复杂的科研设计、数据结构、统计分析和结果解释，都需要运用一定的方法表达或报告，读者才能了解其内容。如何表达、描述和报告，是一个非常值得推敲的问题。在医学论文中，如何陈述统计学处理的内容，如何言简意赅地报告统计分析结果，是常被人们忽视且严重影响论文质量的问题。很多人辨别不清统计资料的类型，导致统计分析方法的误用；很多人将百分比与百分率混为一谈，得出荒谬的结论，很多人误用 $\bar{x} \pm s$ 和 $\bar{x} \pm s_{\bar{x}}$ 表达定量资料，以假乱真；很多人编制的统计表条理不清，数据含义不明，严重地影响了读者的阅读效果；很多人绘制的统计图违反数学原则，呈现出错误的变化趋势。综上所述，不仅在生物医学科研究和论文写作中离不开统计表达与描述方面的知识和技术，而且广大实际工作者迫切需要获得这方面的正确指导，统计学三型理论可以很好地完成此重任。

本书运用统计学三型理论，透过各种具体的统计表达和描述方面问题的“表现型”，揭示其“原型”，进而将“原型”正确地转变为“标准型”，使统计表达与描述方面的问题尽可能得到圆满解决。本书不仅适用于学习和使用统计学学生和工作人员，也适合撰写各种报告和文书的学者。

前言

统计表达与描述属于统计学中的基础内容，但在实际应用中，人们常会犯这样或那样的错误。是人们对基础知识的轻视，还是这些内容中有什么让人难以理解或掌握的东西？导致人们误用的原因可能很复杂，但根本原因可能是人们对统计学基础知识缺乏了解。与实际问题有关的统计学问题通常都存在3种类型，即表现型、原型和标准型。即便是关于“统计资料性质”的划分或表达稍复杂统计资料的统计表和统计图，若只从其表现型入手，不假思索或照猫画虎地就写出其最终结果，没弄清资料要反映的本质，不仅效果不佳，有时会弄巧成拙、杂乱无章，有时甚至会一叶障目、颠倒是非。

本书针对“统计表达与描述”涉及的主要内容，分五篇进行阐述。第一篇基础篇，科学地总结了统计学思想，创造性地提出了统计学三型理论，提纲挈领地概述了统计学内容，深入浅出地介绍了资料类型及其错误辨析与释疑；接着围绕提高学术期刊论文统计学应用质量，阐述了如何修改“稿约”，如何撰写论文中的“统计学处理”部分，最后，还强调了如何报告统计分析结果。第二篇资料表达篇，就各种统计表和统计图的标准模式、非标准模式的修改技巧等非常实用而又很容易出错的内容，做了详细介绍。第三篇资料描述篇，就相对数、平均指标、变异指标、误差等基本统计量和基本概念、常见错误和修改技巧，做了全面而系统的讲解。第四篇基金论文错误辨析篇，对十几种国家级生物医学期刊中发表的基金资助课题写成的论文，仅从统计表达与描述角度进行分析，就揭示出很多统计表达与描述方面的错误，对反映出来的错误进行了辨析与释疑。第五篇概率分布篇，随机

变量的概率分布是统计学的理论基础，也是描述统计资料变化规律的重要技术方法。但这些内容很难学习，掌握起来有一定困难。好在已编写出反映各种概率密度函数曲线图形的 SAS 程序，在这些程序的帮助下，绘制和理解这些随机变量的概率分布规律图就变得容易多了。

在本书的编写过程中，本室周诗国和柳伟伟两位年轻老师承担了比较繁重的编写和审阅任务；本室的博士研究生葛毅和硕士研究生程瑞专也承担了很重要的编写和审阅任务；首都医科大学的刘惠刚副教授、彭迎春和沈艳红讲师等为本书的第一篇部分章节和第四篇撰写作出了突出贡献；本院 2006 级硕士研究生李鹏和王诗鸿，于 2006 年就为本书做了很多基础性工作，书中的很多统计表都曾由他们首次修改过；北京邮电大学已毕业的硕士研究生胡纯严为本书定制了方便快捷调用 SAS 程序的软件，名为“SAS PAL”，为读者提高调用 SAS 程序的准确性和效率贡献了很大的力量。

最后，还要诚挚地感谢人民军医出版社的领导和负责本书的编辑于岚，正是他们的慧眼，才使“统计学三型理论”在生物医学科研这片沃土中生根、发芽、开花、结果。

胡良平

2007-12-13

于北京军事医学科学院生物医学统计学咨询中心

目 录

第一篇 基础篇

第1章 统计学思想	3
一、医学统计学定义综述	3
二、问题的提出	4
三、时代呼唤更新的统计学思想	5
四、统计学思维模式包含的“八性”	5
五、统计学思维模式包含的“八思维”	11
六、小结	16
第2章 统计学三型理论	17
一、统计学三型理论的定义及各型间的相互关系	17
二、统计学三型理论产生的背景	19
三、统计学三型理论提出的依据	21
四、统计学三型理论适用的范围	21
五、统计学三型理论应用的样例	21
六、小结	26
第3章 医学统计学内容概述	27
一、统计研究设计概述	27
二、统计表达与描述概述	33
三、统计分析方法的合理选择与实现	34
四、结果的解释	37
五、结论的陈述	38
六、学习医学统计学的方法概述	38
七、小结	39
第4章 资料类型及其错误辨析与释疑	40
一、资料类型的两种划分方法	40
二、资料类型识别中常犯的错误	42
三、误判资料类型的案例	43

四、小结	46
第 5 章 应写入生物医学期刊“稿约”中的统计学内容	47
一、统计学在医学科研和医学写作中的地位和作用	47
二、在医学科研和医学写作中如何正确运用统计学	48
三、生物医学期刊“稿约”中应增加的内容	49
四、小结	51
第 6 章 如何撰写医学论文中的统计学处理部分	52
一、生物医学论文中统计学部分写法的回顾	52
二、对上述各例错误的辨析与释疑	53
三、论文中统计学处理部分的一般形式及错误辨析	54
四、统计学处理部分究竟应该包括哪些内容	55
五、小结	56
第 7 章 如何报告统计分析结果	57
一、如何报告单变量统计分析结果	57
二、如何报告平均值比较的结果	57
三、如何报告平均秩比较的结果	59
四、如何报告样本频率比较的结果	60
五、如何报告各组患者人数分布规律的分析结果	61
六、如何报告指标之间相关关系的分析结果	61
七、如何报告指标之间依赖关系的分析结果	62
八、小结	62

第二篇 资料表达篇

第 8 章 详细表达多因素设计定量资料的标准模式	65
一、随机区组设计	65
二、含一个协变量的随机区组设计	66
三、平衡不完全随机区组设计	67
四、拉丁方设计	68
五、交叉设计	69
六、无重复实验的双因素设计	72
七、嵌套设计	73
八、裂区设计	75

九、析因设计	78
十、正交设计	83
十一、均匀设计	85
十二、重复测量设计	87
十三、小结	94
第 9 章 结果为定量资料的各种统计表的修改技巧	95
一、多因素析因设计定量资料对应的统计表	95
二、重复测量设计定量资料对应的统计表	101
三、多因素非平衡组合实验定量资料对应的统计表	127
四、特殊情况下定量资料对应的统计表	145
五、小结	152
第 10 章 正确表达各种定性资料的标准模式	153
一、 2×2 表资料	153
二、 $2 \times C$ 表资料和 $R \times 2$ 表资料	156
三、 $R \times C$ 表资料	157
四、高维列联表资料	160
五、具有重复测量设计的定性资料	163
六、小结	164
第 11 章 结果为定性资料的各种统计表的修改技巧	165
一、表中数据含义不清	165
二、重复测量定性资料表达方式的修改技巧	168
三、多种信息混于一表的修改	174
四、对高维列联表的修改	175
五、统计表纵、横标目混乱	177
六、小结	178
第 12 章 混合型资料的各种统计表的修改技巧	179
一、合理编制混合型资料统计表的重要性	179
二、统计表标题冗长	181
三、未合理安排统计表的分组标志	183
四、未妥善安排统计表的标目	184
五、统计表的表头层次过多	186
六、统计表内“注释内容”喧宾夺主	188

七、统计表中资料表达不科学	190
八、纵横标目安排不当且统计表过于庞杂	192
九、统计表中绝对数与相对数交叉出现	193
十、统计表层次太多且夹杂着统计计算	195
十一、统计表中横、竖线条过多且有斜线	197
十二、小结	199
第 13 章 常见统计图的标准模式	200
一、统计图概述	200
二、常见统计图的应用场合及举例	203
三、小结	223
第三篇 资料描述篇	
第 14 章 相对数及其错误辨析与释疑	227
一、相对数的应用场合	227
二、相对数的种类与计算	228
三、比和率的区别和联系	233
四、应用相对数时的注意事项	234
五、相对数应用举例	235
六、相对数应用中常犯的错误及错误辨析	237
七、小结	241
第 15 章 平均与变异指标的应用及其错误辨析与释疑	242
一、平均指标的概念与应用	242
二、变异指标的概念与应用	246
三、参考值范围与置信区间的概念与应用	249
四、平均与变异指标结合应用中常见错误的辨析与释疑	250
五、小结	253
第 16 章 各种误差及其发现和处理方法	254
一、误差的概念	254
二、研究误差的目的与意义	254
三、误差的分类	254
四、误差的数学表达	256
五、系统误差的发现与消除	257

六、随机抽样误差的控制	260
七、过失误差的发现与消除	263
八、有效数字与数字舍入规则	268
九、常见的变量变换方法	273
十、小结	275

第四篇 基金论文错误辨析篇

第 17 章 基金论文中描述性统计错误辨析与释疑 (1)279

一、统计表内表达的内容含糊不清	279
二、纵、横标目颠倒	284
三、同一统计表中表达多项内容	289
四、统计图坐标轴刻度值标注不符合数学原则	292
五、统计图类型选择不当	296
六、小结	299

第 18 章 基金论文中描述性统计错误辨析与释疑 (2)300

一、统计图坐标轴上的刻度值标注不符合数学原则	300
二、误用正态分布法描述呈偏态分布的资料	304
三、计算相对数时分母太小	304
四、用条图表达适合用线图表达的资料	308
五、纵、横标目颠倒	312
六、统计表内表达的内容含混不清	312
七、统计表的编制与实验设计类型不吻合	313
八、统计表纵标目含混不清	314
九、以率代替比	315
十、同一统计表中表达多项内容	316
十一、小结	317

第五篇 概率分布篇

第 19 章 常见离散型随机变量概率分布321

一、离散型随机变量及其常见类型	321
二、二项分布	322
三、Poisson 分布	327

统计学三型理论在统计表达与描述中的应用

四、超几何分布	332
五、应用	336
六、负二项分布	337
七、小结	342
第 20 章 常见连续型随机变量概率分布	343
一、连续型随机变量的概念	343
二、连续型随机变量的概率分布	343
三、正态分布	344
四、 Γ 分布	351
五、 χ^2 分布	353
六、 t 分布	356
七、 F 分布	359
八、指数分布	361
九、威布尔分布	364
十、均匀分布	366
十一、小结	368
附录 A 胡良平专著及配套软件简介	369

第一篇 基础篇

第 1 章 统计学思想

很多人（甚至包括大多数统计学工作者）都认为统计学是一种“计算工具”，其实，这种认识具有很大的片面性。因为需要运用统计学解决的问题中，绝大多数都应从收集资料开始。如何才能获得准确的资料呢？还得从统计研究设计开始。事实上，比统计研究设计更重要的是如何将专业知识与统计学知识相结合，全面考虑问题，使期望达到的研究目的能更好地实现，这就是统计学思想。换句话说，统计学思想就是全面而又系统地考虑问题，设法透过事物的现象去探究其本质。本章将围绕这一主题，结合实例详细阐述统计学思想的内涵。

一、医学统计学定义综述

（一）传统的定义

统计学 (statistics) 与医学统计学 (statistics in medicine, medical statistics, statistics of medicine) 的定义都很多，这里介绍有代表性的几种，供大家参考，以便读者对统计学，特别是医学统计学有一个基本的认识。

定义 1-1: 统计学 (statistics): “A science dealing with the collection, analysis, interpretation, and presentation of masses of numerical data” (Webster's International Dictionary. 摘自: 方积乾. 卫生统计学. 第 5 版. 北京: 人民卫生出版社, 2004)。

定义 1-2: 统计学 (statistics): “The science and art of dealing with variation in data through collection, classification, and analysis in such a way as to obtain reliable results” (John M. Last. A Dictionary of Epidemiology. 摘自: 方积乾. 卫生统计学. 第 5 版. 北京: 人民卫生出版社, 2004)。

定义 1-3: 统计学 (statistics): 对令人困惑费解的问题做出数字设想的艺术 (摘自: 魏宗舒等译. 统计学. 北京: 中国统计出版社, 1997, 原书由美国 David Freedman 等著)。

定义 1-4: 统计学 (statistics): 在相对有限的样本数据上，对特定的随机现象做推断的学科 (摘自: 孙尚拱译. 生物统计学基础. 北京: 科学出版社, 2004. 原书由美国 Bernard Rosner 著)。

定义 1-5: 统计学 (statistics): “Statistics is a collection of procedures and principles for gathering data and analyzing information in order to help people make decision when faced with uncertainty” (摘自: Jessica M Utts, Robert F. Hechard. Mind on Statistics. 22nd ed. 北京: 机械工业出版社, 2005)。

定义 1-6: 医学统计学是运用概率论和数理统计的原理、方法，结合医学实际，研究

数字资料的搜集、整理分析和推断的一门学科（摘自：杨树勤. 中国医学百科全书医学统计学. 上海：上海科学技术出版社，1985）。

定义 1-7: 医学统计学是一门运用统计学的原理和方法，研究医学科研中有关数据的搜集、整理、分析的科学（摘自：孙振球. 医学统计学，北京：人民卫生出版社，2002）。

定义 1-8: 医学统计学是用统计学原理和方法研究生物医学问题的一门学科（摘自：徐勇勇. 医学统计学. 第2版. 北京：高等教育出版社，2004）。

定义 1-9: 统计学（statistics）是关于数据（data）的学问，是从数据中提取信息、知识的一门科学与艺术，包括研究设计、数据搜集、数据整理、数据分析和结果报告等步骤（摘自：方积乾. 生物医学研究的统计方法. 北京：高等教育出版社，2007）。

（二）消化吸收与发扬光大

从上述关于统计学的前8种定义中不难看出，统计学家几乎都认为统计学是收集资料和分析资料的“工具”。事实上，很多资料并非业已存在的，而是需要有科学研究头脑的人根据研究目的，有的放矢地去做出科学完善的设计，在正确设计指导下，进行调查或实验，才能产生出数据（即资料），若等到有了数据后再运用统计学知识去收集资料，可能已经“晚”了。若缺乏正确的统计学思想，就可能做出错误的设计，在错误的设计（有时没有任何设计）之下，产生出来的资料难免丢三落四、鱼目混珠，很难用统计学方法去粗取精、去伪存真，必然导致统计分析结果不可信。因此，对统计学这门学科的认识也应与时俱进，应以辩证、全面、深刻和发展的眼光来审视它。

以笔者肤浅之见：统计学是一门将唯物辩证法引入自然科学，凭借其科学的思维来实现正确的科研设计和处理数据中变异性的科学与艺术，它集“基本常识、专业知识、数学、计算机技术和唯物辩证法”于一身，具体内容包括统计研究设计，收集、表达、分析、解释数据和报告结果，以达到“透过事物的现象看清其本质”的目的，从而为了实现某种研究目的而以较少的人力、物力、财力和时间获得准确可靠的结果，尽可能使所得到的结论经得起时间的考验，并使获得的知识和经验有利于指导未来的科研实践。

二、问题的提出

临床医师不仅要为患者诊治疾病，还要从事医学科研工作。科学研究的目的就是探索人类未知的事物，以便更好地认识世界、认识自然、认识社会，从而获得新的知识去征服世界、改造世界，更好地造福于人类。众所周知，未知事物充满无数的变数，用统计学的语言来描述，就是所研究的问题需要通过考查多因素对多指标的影响情况，并需要运用数学、统计学、计算机和各领域的专业知识，对所研究问题进行综合分析和推理，才有可能得出科学的结论，用于指导未来的各项理论研究和生产实践活动。

在科学研究的全过程中，统计学起着越来越重要的作用。遗憾的是，统计学的理论比较深奥，要考虑的问题非常多，应用时灵活性又非常大，当临床医师和生物医学科研工作者（以下简称实际工作者）要运用较复杂的实验设计知识和统计分析知识时，常感到心中无底，有时甚至望而却步。问题的症结并非出在实际工作者本身，而是出在现行的生物（或

医学，或卫生）统计学教科书。因为现行的统计学教科书中所写的内容几乎都是经过统计学工作者加工后的，它们是一代又一代统计学工作者经过多年“修剪”出来的“**标准型**”。实际工作者拿着自己的问题（通常是披着假象的“**表现型**”），试图通过比照这些“**标准型**”找到解决问题的办法，但“**表现型**”与“**标准型**”之间常常并非完全一致，因此，盲目套用十有八九会出错。要想不出错，必须弄清每个“**表现型**”的本质（即问题的“**原型**”），倘若能将“**原型**”正确地转变成对应的“**标准型**”，问题也就迎刃而解了。

三、时代呼唤更新的统计学思想

（一）时代为何要呼唤更新的统计学思想

把复杂的统计计算视为学习统计学的难点，在计算机技术十分落后的年代是可以被接受的；把统计学中的每个内容都像修饰后的“模特”那样，使其成为“**标准型**”，这在高等院校还很少开设统计学课程的年代是必需的。但几十年，一直沿用这种思想，去编写统计学教科书、去实施统计学教学，不仅浪费了无数学习者和使用者的时间和精力，而且束缚了他们的想像力和创造力。一个不可否认的事实是，一代又一代曾学过生物医学统计学的人们，他们中的许多人学了多遍统计学仍不得要领。这样的统计学教科书，是不受实际科研工作者欢迎的！一本好的统计学教科书应该明确讲述与每个实际问题对应的前述“三型”（绝对不能只讲述“**标准型**”），实际工作者才可以少花时间和精力（因为他们不可能像统计学工作者那样，专攻统计学！），学到统计学的精髓，才能在生物医学科研工作中发挥统计学的作用，真正达到提高统计学的应用水平和提高科研质量的目的。

（二）何谓更新的统计学思想

科学技术日新月异地发展，对统计学提出了越来越高的要求。时代呼唤更新的统计学思想，而不是照抄或盲目套用那些繁琐的统计学公式。新的统计学思想是什么？是用辩证的思维去观察事物，是用透视的眼光去洞察事物，是透过现象看本质的统计学思维模式，可概括为“八性”和“八思维”。

统计学在看待事物和处理问题时，离不开下列“八性”，即**系统性与代表性、随机性与均衡性、概括性与延展性和自悖性与相合性**。同时，还离不开下列“八思维”，要**弱化静态思维强化动态思维**；要**突破正向思维，巧用逆向思维**；要**跳出简单思维，步入复杂思维**；要**活用横向思维，发展纵向思维**。

四、统计学思维模式包含的“八性”

（一）统计学中的系统性

统计学具有系统性特点。统计学强调考虑问题应当系统而全面，决不当坐井观天、妄下结论。而许多实际工作者在科研工作中经常顾此失彼、丢三落四，常因考虑问题不系