



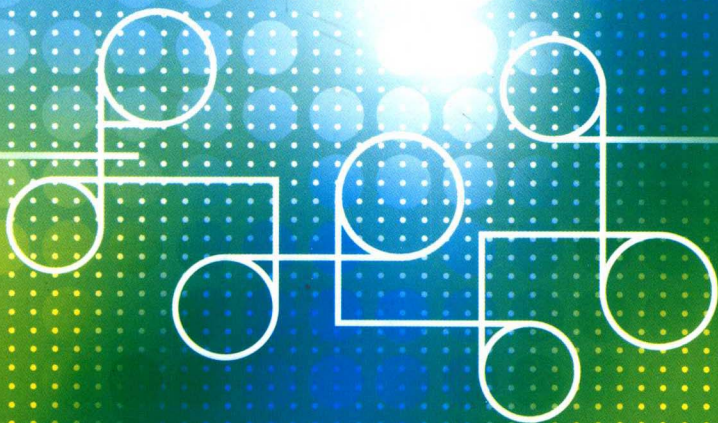
全国统计教材编审委员会“十二五”规划教材

江苏省高等学校精品教材

医学统计学

第三版

◎主编 / 于浩



 中国统计出版社
China Statistics Press



全国统计教材编审委员会“十二五”规划教材

江苏省高等学校精品教材

医学统计学

第三版

主编 于 浩
编者 于 浩 吕 军 陈 峰 易洪刚
赵 杨 姜冬华 柏建岭 荀鹏程

图书在版编目(CIP)数据

医学统计学 / 于浩主编. — 3 版. — 北京: 中国统计出版社, 2013. 3

全国统计教材编审委员会“十二五”规划教材 江苏省高等学校精品教材

ISBN 978-7-5037-6791-3

I. ①医… II. ①于… III. ①医学统计—高等学校—教材 IV. ①R195.1

中国版本图书馆 CIP 数据核字(2013)第 048941 号

医学统计学

作 者/于 浩

责任编辑/徐 颖

装帧设计/上智博文

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://csp.stats.gov.cn>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/710×1000mm 1/16

字 数/225 字

印 张/12.75

印 数/1—4000 册

版 别/2013 年 7 月第 3 版

版 次/2013 年 7 月第 1 次印刷

定 价/29.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

国家统计局

全国统计教材编审委员会

顾 问 罗 兰 袁 卫 冯士雍 吴喜之
 方积乾 王吉利 庞 皓 李子奈

主 任 徐一帆

副主任 严建辉 田鲁生 邱 东 施建军
 耿 直 徐勇勇

委 员 (按姓氏笔划排序)

丁立宏	万崇华	马 骏	毛有丰	王兆军
王佐仁	王振龙	王惠文	丘京南	史代敏
龙 玲	刘建平	刘俊昌	向书坚	孙秋碧
朱 胜	朱仲义	许 鹏	余华银	张小斐
张仲梁	张忠占	李 康	李兴绪	李宝瑜
李金昌	李朝鲜	杨 虎	杨沛华	杨映霜
汪荣明	肖红叶	苏为华	陈 峰	陈相成
房祥忠	林金官	罗良清	郑 明	柯惠新
柳 青	胡太忠	贺 佳	赵彦云	赵耐青
凌 亢	唐年胜	徐天和	徐国祥	郭建华
崔恒建	傅德印	景学安	曾五一	程维虎
蒋 萍	潘 璠	颜 虹		

出版说明

全国统计教材编审委员会是国家统计局领导下的、全国统计教材建设工作的最高指导机构和咨询机构,自1988年成立以来,分别组织编写和出版了“七五”至“十一五”全国统计规划教材。

“十二五”时期,是我国全面实施素质教育,全面提高高等教育质量,深化教育体制改革,推动教育事业科学发展,提高教育现代化水平的时期。“十二五”伊始,统计学迎来了历史性的重大变革和飞跃。2011年2月,在国务院学位委员会第28次会议通过的新的《学位授予和人才培养学科目录(2011)》(以下简称“学科目录”)中,统计学从数学和经济学中独立出来,成为一级学科。这一变革和飞跃将对中国统计教育事业产生巨大而深远的影响,中国统计教育事业将在“十二五”时期发生积极变化。

正是在这一背景下,全国统计教材编审委员会制定了《“十二五”全国统计教材建设规划》(以下简称“规划”)。根据“学科目录”在统计学下设有数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科的构架,“规划”对“十二五”全国统计规划教材建设作了全面部署,具有以下特点:

第一,打破以往统计规划教材出版学科单一的格局。全面发展数理统计学,社会经济统计学,生物卫生统计学,金融统计、风险管理与精算学,应用统计5个二级学科规划教材的出版,使“十二五”全国统计规划教材涵盖5个二级学科,形成学科全面并平衡发展的出版局面。

第二,打破以往统计规划教材出版层次单一的格局。在编写出版好各学科本科生教材的基础上,对研究生教材出版进行深入研究,出版一批高水平高层次的研究生教材,为我国研究生教育、尤其是应用统计研究生教育提供教学服务。同时,积极重视统计专科教材出版,联合各专科院校,组织编写和出版适应统计专科教学和学习的优秀教材。

第三,打破以往统计规划教材出版品种单一的格局。鼓励内容创新,联系统计实践,具有教学内容和教学方法特色的、各高校自编的相同内容选题的精品教材出版,促进统计教学向创新性、创造性和多样性发展。

第四,重视非统计专业的统计教材出版。探讨对非统计专业学生的统计教学问题,为非统计专业学生组织编写和出版概念准确、叙述简练、深入浅出、表达方式活泼、练习题贴近社会生活的统计教材,使统计思想和统计理念深入非统计专业学生,以达到统计教学的最大效果。

第五,重视配合教师教学使用的电子课件和辅助学生学习使用的电子产品的配套出版,促进高校统计教学电子化建设,以期最后能形成系统,提高统计教育现代化水平。

第六,重视对已经出版的统计规划教材的培育和提高,本着去粗存精、去旧加新、与时俱进的原则,继续优化已经出版的统计教材的内容和写作,强化配套课件和习题解答,使它们成为精品,最后锤炼成为经典。

“十二五”期间,编审委员会将本着“重质量,求创新,出精品,育经典”的宗旨,组织我国统计教育界专家学者,编写和编辑出版好本轮教材。本轮教材出版后,将能够形成学科齐全、层次分明、品种多样、配套系统的高质量立体式结构,使我国统计规划教材建设再上新台阶,这将对推动我国统计教育和统计教材改革,推动我国统计教育事业科学发展,提高我国统计教育现代化水平产生积极意义。

让教师的教学和学生的学习事半功倍,并使学生在毕业之后能够学以致用,是本轮教材的追求。编审委员会将努力使本轮教材好教、好学、好用,尽力使它们在内容上和形式上都向国外先进统计教材看齐。限于水平和经验,在教材的编写和编辑出版过程中仍有不足,恳请广大师生和社会读者提出批评和建议,我们将虚心接受,并诚挚感谢!

国家统计局
全国统计教材编审委员会
2012年7月

第三版前言

本书介绍了医学统计学的基本概念、基本理论和基本方法,是医学生《医学统计学》课程的入门教材。全书共 10 章,第 1 章绪论,介绍统计学的一些基本概念;第 2 章介绍统计资料的整理与描述;第 3 章介绍了统计学中最重要的两个理论分布:正态分布和二项分布,是统计学的基础;第 4 章以均数为例介绍了统计推断的基本概念、内容和方法,是全书的核心章节;第 5、第 6、第 8 章进一步介绍了不同资料的统计推断方法;第 7 章介绍了两个变量间线性关系的描述与分析;第 9 章介绍了研究设计的基本概念和基本方法;第 10 章是综合分析,旨在通过实例引导读者综合运用所学的方法对实际资料进行分析,在书本知识和实际问题之间架起一座桥梁,从而提高分析和解决实际问题的能力,这是本书的特色。本书语言精练、篇幅短小,适合 48 学时以内的教学。

教学中发现,虽然学生学会了用统计学方法分析研究数据,却不会将得到的统计学结果与专业结果融合,尤其是在撰写专业论文时,不知道如何根据统计学结果下专业结论。因此,本次修订重点在如何根据统计分析结果表述专业上的研究结论,每个例题后均增加了专业结论。另外,为加深对统计学概念的理解和方法的应用,本次修订特意增加了 50 道选择题,供读者课后练习。

本书的编写和出版自始至终得到了江苏省教育厅、南京医科大学的关心和资助,以及中国统计出版社的大力支持。第一版于 2004 年 12 月出版,被列入江苏省精品教材建设规划;2005 年被遴

选为江苏省高等学校精品教材,并于2005年12月再版;2008年1月由台湾合记圖書出版社在台湾出版发行;2012年又被全国统计教材编审委员会遴选为全国统计教材编审委员会“十二五”规划教材。

出版9年来一直得到广大读者的关心和厚爱,作者表示衷心感谢。由于编者学识水平有限,难免有不妥之处,敬请广大师生提出宝贵意见。

编 者

2012年冬于南京

第二版前言

《医学统计学》是医学科学的一个组成部分,是医学院校各专业学生的公共必修课。作为 20 世纪发展最快的学科之一,统计学已经应用到几乎所有的学科领域,医学也不例外。最近的一次调查表明,有 76% 的医学研究论文中运用了统计分析方法,但几乎所有的研究论文都运用了统计学思想。医学统计学已经成为 21 世纪医学科学工作者必须掌握的工具之一。

本书介绍了医学统计学的基本概念和方法。全书共 10 章,第 1 章是绪论,介绍一些基本概念;第 2 章介绍统计资料的整理与描述;第 3 章介绍了统计学中最重要的两个理论分布:正态分布和二项分布,是统计学的基础;第 4 章以均数的推断为例介绍了统计推断的基本概念、内容和方法,是全书的核心章节;第 5、第 6、第 8 章进一步介绍了不同资料的统计推断方法;第 7 章介绍了两个变量间线性关系的描述与分析;第 9 章介绍了研究设计的基本概念和基本方法;第 10 章是综合分析,旨在通过实例引导读者综合运用所学方法对实际资料进行分析,在书本知识和实际问题之间架起一座桥梁,从而提高解决实际问题的能力,这是本书的特色。本书可供非预防医学专业的学生使用。

现代教育技术手段的不断发展,使得医学统计学的教学方式也发生了巨大的改变。我们开设的《医学统计学》作为在建的精品课程,已经在网上运行了,网络课程中除了课堂所讲述的内容外,还为学有余力的同学提供了大量的进一步学习的素材和参考资料,其中包括一些英文原版教材、名词解释、统计幽默、统计软件操作指导等,还介绍了国内外一些著名的统计学家。教师通过网络布置作业,自动实时批改作业;

教师和同学们可以在网络课程中自由讨论、交流,加强了师生的沟通,提高了学习效率,不同层次的同学均可有所收益。

医学统计学中的一些概念比较抽象,所涉及的数理统计学理论也比较复杂,为了使这些概念和理论易于理解,便于接受,我们自行研制开发了《医学统计学中随机现象的计算机模拟系统》,采用计算机随机试验的方式,通过动画、图片和文字,验证统计学中的一些基本定理和抽象理论,将抽象的概念具体化、形象化,复杂的理论简单化、生动化。通过教师在课堂演示和学生自己操作模拟系统,加深对基本概念、基础理论的理解,从而架起了抽象理论与实际现象之间的桥梁。实践证明,这种教学方法提高了教学质量,也激发了同学们的学习兴趣。

本书的编写自始至终得到了南京医科大学公共卫生学院的关心和帮助,得到南京医科大学教材建设委员会的大力支持,此次再版得到了江苏省教育厅的资助。柏建岭验算了书中所有实例。作者表示衷心感谢。

由于我们学识水平有限,难免有不妥之处,敬请广大师生提出宝贵意见。

编 者

2005 年冬于南京

目 录

第 1 章 绪论	(1)
1.1 引言	(1)
1.2 几个基本概念	(3)
第 2 章 统计资料的整理与描述	(9)
2.1 频数表	(9)
2.2 描述集中位置的指标	(12)
2.3 描述离散趋势的指标	(14)
2.4 分类资料的率和比	(18)
复习思考题	(24)
第 3 章 正态分布和二项分布	(26)
3.1 随机变量的概率分布	(26)
3.2 正态分布	(27)
3.3 二项分布	(33)
复习思考题	(37)
第 4 章 统计推断基础	(38)
4.1 抽样误差与标准误	(38)
4.2 参数估计	(44)
4.3 假设检验的基本思想与步骤	(48)
4.4 均数比较的假设检验	(51)
4.5 第一类错误与第二类错误	(56)
4.6 假设检验时应注意的问题	(58)
复习思考题	(61)
第 5 章 方差分析	(63)
5.1 方差分析的基本思想	(63)
5.2 完全随机设计资料的方差分析	(66)

5.3	随机区组设计的方差分析	(67)
5.4	多个样本均数间的两两比较	(70)
5.5	方差分析的应用条件	(72)
	复习思考题	(75)
第6章	分类资料的统计推断	(77)
6.1	率的区间估计	(77)
6.2	样本率与总体率的比较	(78)
6.3	两样本率的比较	(80)
6.4	多组率或构成比的比较	(85)
6.5	配对两分类资料的假设检验	(87)
	复习思考题	(88)
第7章	直线相关与回归	(91)
7.1	确定性关系与非确定性关系	(91)
7.2	直线相关	(91)
7.3	直线回归	(96)
7.4	直线相关与回归应用注意事项	(104)
	复习思考题	(106)
第8章	秩和检验	(108)
8.1	秩次与秩和	(108)
8.2	配对设计差值比较的符号秩和检验	(109)
8.3	成组设计两样本比较的秩和检验	(111)
8.4	成组设计多样本比较的秩和检验	(114)
8.5	随机区组设计资料的秩和检验	(117)
8.6	多样本资料的两两比较	(118)
	复习思考题	(120)
第9章	研究设计基础	(123)
9.1	研究设计的意义	(123)
9.2	实验研究的特点	(124)
9.3	实验研究的基本要素	(124)
9.4	实验设计的基本原则	(127)
9.5	研究设计的常见类型	(131)

9.6 常见的抽样方法	(134)
复习思考题	(137)
第 10 章 统计方法的综合运用	(138)
10.1 统计学设计及统计方法的选择	(138)
10.2 基本统计方法选择的流程图	(151)
附录 A 统计用表	(154)
附表 1 标准正态分布曲线下的面积, $\Phi(-u)$ 值	(154)
附表 2 t 界值表	(155)
附表 3 F 界值表(方差分析用)	(156)
附表 4 F 界值表(方差齐性检验用)	(160)
附表 5 q 界值表(Student-Newman-Keuls 法)	(162)
附表 6 q' 界值表(Dunnett 法)	(163)
附表 7 百分率的可信区间	(164)
附表 8 χ^2 界值表	(170)
附表 9 T 界值表(配对比较的符号秩和检验用)	(171)
附表 10 T 界值表(两组比较的秩和检验用)	(172)
附表 11 H 界值表(三组比较的秩和检验 Kruskal-Wallis 法)	(173)
附表 12 M 界值表(配伍组设计的秩和检验 Friedman 法)	(174)
附表 13 随机排列表($n=20$)	(175)
附表 14 随机数字表	(176)
附录 B 英汉医学统计学词汇	(177)
附录 C 选择题	(185)
参考文献	(190)

遵循统计学设计的基本原理,对实验的每个环节进行周密设计,目的在于创造一致的对比条件,有效地控制实验误差,以较少的人力、物力和时间取得较好的效果。详见第9章。

2) 分布理论。是统计学的理论基础,主要用于探讨疾病的统计分布规律,为选择相应的统计分析方法(如假设检验,统计建模,质量控制,疾病监测方法等)提供依据,是制订正常值范围,研究疾病等在空间上、时间上或人群中的分布规律的重要手段。详见第3章。

3) 统计描述。对原始资料进行一般性的描述,以期得到初步的了解和直观印象。如平均水平,离散程度,分布形状等。可用文字、统计指标或统计图表表示。详见第2章。

4) 参数估计与假设检验。是推断统计学的重要组成部分。在大多数医学科研中需要对研究对象的全体(称之为总体)的某些特征参数作出适当的估计。包括均数、百分率等;或对资料是否来自具有某种属性的总体进行检验。常用于新药鉴定、病因分析、理化检验方法和技术水平的考核等。包括 t 检验、方差分析、 χ^2 检验、秩和检验等。详见第4章、第5章、第6章、第8章。

5) 相关与回归。主要研究两变量之间的关系,常用于病因学研究、发育或生理功能评价以及各种预测、趋势分析等。包括线性相关、直线回归。详见第7章。

作者专门安排了“统计方法的综合运用”一章,以帮助同学们正确理解和灵活运用所学医学统计学知识,在书本知识和实际问题之间架起一座桥梁,从而提高分析问题和解决问题的能力。

1.1.2 学习医学统计学的目的与要求

医学生学习医学统计学并非要使其成为统计专业人员,其目的在于使大家具备新的推理思维,学会从不确定性和概率的角度去考虑问题;学会结合专业问题合理进行课题设计,通过精细的研究获得可靠、准确的资料;学会正确运用统计方法充分挖掘资料中隐含的信息,并能恰如其分地作出理性概括,写成具有一定学术水平的研究报告或科学论文,提高自身的科研素养。

为此,医药卫生各专业的学生必须学习医学统计学。

1.1.3 如何学好医学统计学

统计学的思维是用变异与不确定性,机遇与概率的观点去考虑问题,在齐同的基础上进行比较、分析,依据概率用逻辑推理去作结论,属于从个别到一般的归纳推理型思维。这在一定程度上与人们在其他学科学习和日常生活中养成的确定性的、偏于从一般到个别的演绎推理型的思维方法有所不同,初学统计应注意这

一点。

统计离不开数字,每个数字都有其实际意义。表面上看起来杂乱无章的数字,其间却隐含着内在的规律。因此不要厌烦数字,应重视原始数据的完整性和准确性,对数据处理持严肃、认真、实事求是的科学态度,严禁伪造和篡改统计数字。

统计亦离不开公式和计算。统计学中的公式都是由实际问题引申出来的,一般都有其实际意义,虽不要求掌握其数学推导,但了解其直观意义、用途和应用条件是必要的,学习时要留心有关解释,并多加思考,这将有助于对公式的理解和正确应用。学习医学统计学还应该多做练习,本书的每一章均配有一定数量的习题,通过做练习,帮助大家学会思考,熟悉概念,学会正确运用统计方法处理实际问题。统计中遇到的计算无非就是加、减、乘、除、平方、开方、对数、指数,再加上查表等,并不复杂,尽管现在有很多统计分析软件包,可以省去繁琐的计算,但如果对统计概念理解不透,统计方法选择不当,对统计分析软件输出的结果亦不会有更深刻的认识和正确的解释。因此,做一些简单的、数据量少的练习是必要的,只有这样才能加深对书本知识的理解,体会出其中滋味。

正确应用统计方法,能帮助我们正确认识客观事物,阐明事物的固有规律,从而把感性认识提升到理性认识。但统计不是万能的,它绝不能改变事物的本来面目,把不存在的规律“创造”出来。有些人在进行试验之前没有充分考虑,收集了一些不准确、不可靠或不全面的资料,希望用统计方法来弥补,这是不可能的,统计只能认识规律而不能“创造”规律。

最后必须注意,统计分析手段需要有正确的医学理论作指导,不能将医学问题归结到纯粹的数量问题,否则会归纳出错误的,甚至是荒谬的结论。统计学上所得到的结论都具有概率性,它不能证明什么,但可提高你的分辨能力和判断能力,为科学决策提供依据。

1.2 几个基本概念

1.2.1 同质

性质相同的事物称为同质(homogeneity)的,否则称为异质的或间杂(heterogeneity)的。观察单位间的同质性是进行研究的前提,也是统计分析的必备条件,缺乏同质性的观察单位是不能笼统地混在一起进行分析的。如不同年龄组男童的身高不能计算平均数,因为所得结果没有意义。

不同研究中或同一研究中不同观察指标对观察对象的同质性的要求不同,即同质是相对的。例如,男性身高与女性身高有着本质的差别,因此,在考虑身高这

一指标时,不能把不同性别的人混在一起,此时,不同性别表示不同质;而在研究白细胞计数这一指标时,因性别对该指标影响甚微,故可以把不同性别的人放在一起分析。又如,在某新药的临床试验中,计算有效率的观察病例必须患同一疾病,甚至具有相同的病型、病情、病程等,对同质性的要求是很严格的;而计算不良反应发生率,通常可将不同病种的病例合起来统计,此时对同质性的要求只有一条:按规定服用该新药。

1.2.2 变异

宇宙中的事物千差万别,各不相同,即使是同质事物,就某一观察指标来看,各观察单位(亦称个体)之间也有差别,这种同质事物间的差别称为变异(variation)。例如,研究儿童的身体发育,同性别、同年龄儿童的身高,有高有矮,各不相同,称为身高的变异。由于观察单位通常即观察个体,故变异亦称个体变异(individual variation)。变异表现在两个方面:其一,个体与个体间的差别;其二,同一个体重复测量值间的差别。变异是宇宙事物的个性反映,在生物学和医学现象中尤为明显。

变异是由于一种或多种不可控因素(已知的和未知的)以不同程度、不同形式作用于生物体的综合表现。如果我们掌握了所有因素对生物体的作用机制,那么,生物体的某指标之观察值就是可预测的了。有些指标的变异原因已被人们认识,例如,染色体决定了新生儿的性别;有些指标的变异原因已被认识一部分,比如,人的身高受遗传和后天营养的影响,但尚有一部分影响因素是未知的;更多的情况下,影响变异的因素是未知的。就每个观察单位而言,其观察指标的变异是不可预测的,或者说是随机的。观察指标的表述用变量(variable),或称随机变量(random variable)。当观察值的个数达到足够多时,其分布将趋于稳定,并最终服从于总体分布(distribution of population)。

个体变异现象广泛存在于人体及其他生物体,是个性的反映。虽然每个个体的变异表现出一定的随机性和不可预测性,但变异并不等于杂乱无章,指标的变异是有规律的,当所观察的个体数足够多时,观察值的分布将呈现一定的规律性,这是总体的反映。统计学就是探讨变异规律并运用其规律性进行深入分析的一门学科。可以这么说,没有变异就没有统计学。

1.2.3 总体、个体和样本

就某一观察指标而言,总体(population)是根据研究目的所确定的同质观察单位的全体;个体(individual)是构成总体的最基本的观察单位;样本(sample)是从总体中随机抽取的部分个体;样本中所包含的个体数称为样本含量(sample size)。

例如,调查某地某年正常成年男子的红细胞计数,则观察对象是该地的正常成年男子,全部正常男子的红细胞计数构成了研究总体,其同质基础是同一地区,同一年份,同为正常人,同为成年男性。观察单位是该地该年的每一个正常成年男子。今从中抽取了20名,测得其红细胞计数,则这是一个样本含量为20的样本。这里的总体只包括(确定的时间、空间范围内)有限个观察单位,称为有限总体(finite population)。有时总体是假想的,如研究某种辅助疗法对肾移植病人生存时间的影响,这里总体的同质基础是同为肾移植病人,同用某种辅助疗法,总体包括设想用该辅助疗法的所有肾移植病人,是没有时间和空间概念的,因而观察单位是无限的,称为无限总体(infinite population)。

医学研究中的总体很多是无限总体,要直接研究总体是不可能的。即使是有限总体,如果包含的观察单位过多,也要花费大量的人力、物力、财力,有时也是不可能的和不必要的。如检查乙肝疫苗的合格率,不可能将所有的疫苗打开逐一检查。所以实际工作中总是从总体中随机抽取一定含量的样本,目的是根据样本所提供的信息推断总体的特征,这是统计推断的根本内容。

1.2.4 随机

随机(random)即机会均等,是为了保证样本对总体的代表性、可靠性,使各对比组间在大量不可控制的非处理因素的分布方面尽量保持均衡一致,而采取的一种统计学措施。随机包含三个方面:

1) 抽样随机。每一个符合条件的实验对象参加实验的机会相同,即总体中每个个体有相同的会被抽到样本中来。抽样随机是保证所得到的样本具有代表性,以使研究所得结论具有普遍意义。

2) 分组随机。每个实验对象分配到同一处理组的机会相同。分组随机是保证各处理组间实验对象尽可能均衡一致,以提高各组间的可比性。

3) 实验顺序随机。每个实验对象先后接受处理的机会相同;实验顺序的随机就是平衡实验顺序对观察结果的可能影响。

随机是实验设计中必须遵循的基本原则之一,在实验对象的抽样、分组、实施过程中均应遵循随机化原则。随机化方法很多,普通用的抽签法或掷骰子法就是最原始、最简单的随机化方法。在科学实验中随机化是通过计算机产生的随机数(random number)实现的。详见第9章。

1.2.5 变量的分类

统计分析最基本的是变量,即观察对象个体的特征或测量的结果。由于个体的特征或指标存在个体差异,观察结果在测量前不能准确预测,故称为随机变量