



全国统计教材编审委员会“十二五”规划教材

医学统计学基础

新编

◎陈长生 主编

 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

医学统计学基础

新 编

◎陈长生 主编



中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

医学统计学基础 / 陈长生主编. — 北京 : 中国统计出版社, 2014.10

全国统计教材编审委员会“十二五”规划教材

ISBN 978—7—5037—7329—7

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

中国版本图书馆 CIP 数据核字(2014)第 240322 号

医学统计学基础

作 者/陈长生

责任编辑/梁 超

封面设计/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/三河市利兴印刷有限公司

经 销/新华书店

开 本/710×1000mm 1/16

字 数/430 千字

印 张/23.5

版 别/2014 年 11 月第 1 版

版 次/2014 年 11 月第 1 次印刷

定 价/46.00 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		

《医学统计学基础》编委会

主 编 陈长生

副 主 编 王 玖 王 霞

编 委 (按姓氏笔划排序)

万 毅 王 玖 王 娟 王 霞

王小丹 王素珍 田 冶 石福艳

刘莲花 孙红卫 毕育学 张晋昕

李 强 李长平 陈长生 崔 壮

曹文君 彭志行 颜 艳

学术秘书 万 毅 孙红卫

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用的统计教材,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会有不足,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

前 言

本教材是在全国统计教材编审委员会指导下,从所有医学院校卫生(生物/医学)统计专业课程中遴选出来的首批教学效果好、符合统计学人才培养需求的优秀课程教材。现作为“十二五”全国统计教材国家级规划教材由中国统计出版社出版。

统计学是所有医学院校普遍开设的一门重要的必修基础课程,尤其对于卫生(生物/医学)统计专业与预防医学、卫生管理类专业来说,统计学是一门专业基础课程,目的是让学生掌握医学与卫生研究及工作实践中统计数据的收集、整理和分析的基本原理及应用方法,以便提高统计理论知识水平与实际应用能力。

本书内容包括绪论、计量资料的统计描述、计数资料的统计描述、参数估计、假设检验、 t 检验、 χ^2 检验、多个样本均数比较的方差分析、基于秩转换的非参数检验、两变量回归、线性相关、残差分析,共 12 章。此外,书末附有 4 个附录,其中有各种统计用表 14 种、常用统计图表、R 统计软件、Excel 统计功能简介以及供读者进一步学习的主要参考文献等。

作为教学的扩充内容,免费的 R 统计软件与普通的办公软件 Excel 统计功能简介可指导学生课外阅读,软件的操作应用可以培养学生对统计学方法的综合分析、运用及动手能力,使他们对统计学理论知识的理解更加深刻,处理统计学方法应用问题更加灵活。

本教材的主要教学对象是医学院校四年制本科卫生(生物/医学)统计学专业学生,也适用于预防医学(公共卫生)专业的学生,并可作为五年制以上临床医学专业学生、医学硕士研究生的基础医学统计学教材。

本教材的编写得到中国统计出版社、第四军医大学、广州医学院和滨州医学院领导和有关部门的大力支持。本教材的编委会由国内

11 所高等医学院校的年富力强的统计学教师组成,他们多年从事统计学的教学与科研工作,具有丰富的教学经验和广博深厚的统计理论知识,他们的积极参与和指导是本书具有科学性、理论性和实践性的保证。

2013 年 1 月在西安召开的全国统计教材编审委员会生物卫生统计组“十二五”规划教材研讨会上,来自中山大学公共卫生学院、西安交通大学医学部和南京医科大学公共卫生学院等 12 所医学院校的统计教师对本教材提出了许多宝贵意见。第四军医大学军事预防医学院徐勇勇教授、中山大学公共卫生学院方积乾教授、南京医科大学公共卫生学院陈峰教授、天津医科大学公共卫生学院马骏教授、泰山医学院基础医学院景学安教授与滨州医学院徐天和教授等作为本教材的特邀征求意见专家,始终关心着本教材的编写,他们的意见和建议使本教材大为增色。

感谢所有参加编写的老师为本教材付出的辛勤劳动。共同的劳动和创作使我们增进了友谊,教材的字里行间里也留下了一段段美好的记忆。

由于本人能力所限,教材中肯定存在许多不足,诚恳希望广大师生提出进一步修改意见。

陈长生

2014 年 10 月于西安

目 录

第一章 绪 论	(1)
§ 1.1 统计学与医学	(1)
§ 1.2 统计学的几个基本概念	(2)
§ 1.3 统计学的发展简史	(7)
练习题	(9)
第二章 计量资料的统计描述	(11)
§ 2.1 频数分布	(11)
§ 2.2 集中趋势的描述	(14)
§ 2.3 离散趋势的描述	(18)
§ 2.4 正态分布	(21)
§ 2.5 医学参考值范围的制定	(24)
练习题	(26)
第三章 定性资料的统计描述	(28)
§ 3.1 常用相对数	(28)
§ 3.2 率的标准化法	(33)
§ 3.3 随机变量的概率分布	(39)
练习题	(48)
第四章 参数估计	(51)
§ 4.1 抽样分布	(51)
§ 4.2 抽样误差	(57)
§ 4.3 总体均数的置信区间	(62)
§ 4.4 总体率的置信区间	(67)
练习题	(69)
第五章 假设检验	(71)
§ 5.1 假设检验的基本概念	(71)

§ 5.2	假设检验的基本步骤	(72)
§ 5.3	Z 检验	(75)
§ 5.4	假设检验与区间估计的关系	(81)
§ 5.5	检验效能与样本含量估计	(82)
§ 5.6	正态性检验	(89)
§ 5.7	方差齐性检验	(94)
§ 5.8	变量变换	(96)
§ 5.9	假设检验的正确应用	(97)
练习题		(99)
第六章 t 检验		(101)
§ 6.1	单样本 t 检验	(101)
§ 6.2	配对样本 t 检验	(103)
§ 6.3	两样本 t 检验	(106)
练习题		(110)
第七章 χ^2 检验		(112)
§ 7.1	χ^2 统计量的极限定理	(112)
§ 7.2	四格表资料的 χ^2 检验	(115)
§ 7.3	配对四格表资料的 χ^2 检验	(123)
§ 7.4	四格表资料的精确概率法	(125)
§ 7.5	行×列表资料的 χ^2 检验	(127)
§ 7.6	多个样本率间的多重比较	(133)
§ 7.7	行×列表 χ^2 检验的注意事项	(136)
§ 7.8	有序分组资料的线性趋势检验	(137)
§ 7.9	频数分布拟合优度的 χ^2 检验	(138)
练习题		(140)
第八章 多个样本均数比较的方差分析		(146)
§ 8.1	单因素方差分析	(146)
§ 8.2	方差分析的基本思想及其应用条件	(149)
§ 8.3	完全随机设计资料的方差分析	(152)
§ 8.4	随机区组设计资料的方差分析	(154)
§ 8.5	拉丁方设计资料的方差分析	(158)
§ 8.6	两阶段交叉设计资料的方差分析	(163)

§ 8.7 多个样本均数间的多重比较	(166)
练习题	(171)
第九章 基于秩转换的非参数检验	(174)
§ 9.1 配对样本比较的 Wilcoxon 符号秩检验	(175)
§ 9.2 两个独立样本比较的 Wilcoxon 秩和检验	(182)
§ 9.3 完全随机设计多个样本比较的 Kruskal-Wallis H 检验	(188)
§ 9.4 随机区组设计多个样本比较的 Friedman M 检验	(195)
练习题	(200)
第十章 两变量回归	(204)
§ 10.1 线性回归的概念	(204)
§ 10.2 线性回归方程的建立	(205)
§ 10.3 线性回归分析中的统计推断	(214)
§ 10.4 曲线拟合	(226)
练习题	(233)
第十一章 线性相关	(236)
§ 11.1 线性相关的概念	(236)
§ 11.2 相关系数的计算和意义	(237)
§ 11.3 相关系数的统计推断	(239)
§ 11.4 线性相关与回归的注意事项	(241)
§ 11.5 秩相关	(244)
练习题	(247)
第十二章 残差分析	(249)
§ 12.1 概述	(249)
§ 12.2 异常值检验	(252)
§ 12.3 拟合优度检验	(255)
§ 12.4 一般线性模型的诊断	(256)
练习题	(262)
附录一 统计用表	
附表 1 标准正态分布表($\Phi(z)$ 值, $z \leq 0$)	(263)
附表 2 t 界值表	(264)

附表 3	F 界值表(方差齐性检验用, 双侧界值)	(266)
附表 4	F 界值表(方差分析用, 单侧界值)	(267)
附表 5	q 界值表(Student-Newman-Keuls 法用)	(275)
附表 6 (1)	Dunnett- t 检验界值表(单侧)	(277)
附表 6 (2)	Dunnett- t 检验界值表(双侧)	(279)
附表 7	百分率的可信区间	(281)
附表 8	χ^2 界值表	(287)
附表 9	T 界值表(配对比较的符号秩检验用)	(288)
附表 10	T 界值表(两独立样本比较的秩和检验用)	(290)
附表 11	H 界值表(三组独立样本比较的秩和检验用)	(292)
附表 12	M 界值表(随机区组比较的 Friedman M 检验用)	(293)
附表 13	相关系数 r 界值表	(294)
附表 14	Spearman 秩相关系数 r_s 界值表	(295)
附录二	常用统计图表	(297)
第一节	统计表	(297)
第二节	统计图	(299)
附录三	统计学方法实现——R 统计软件	(305)
第一节	R 软件基础知识	(306)
第二节	对象与数据类型	(308)
第三节	数据读取与整理	(315)
第四节	程序编写	(323)
第五节	R 基本统计方法	(328)
附录四	Excel 统计功能简介	(347)
第一节	Excel 基本知识	(347)
第二节	Excel 的函数	(353)
第三节	创建图表	(358)
第四节	数据统计分析工具	(360)
参考文献		(363)

第一章 绪 论

§ 1.1 统计学与医学

提起统计,很容易让人想到数据汇总,如日常生活中的统计一下医院门诊患者人数、学校各班学生人数等。人类最初的统计确实主要是数据汇总,《史记》上就有“禹平水土,定九州,计民数”的记载。现代的统计是一种对客观现象数量方面进行的调查研究活动,是收集、整理、分析、推断、判断等认识活动的总称,数据汇总仅仅是现代统计工作的一小部分内容。统计方法是现代科学方法之一,统计学(statistics)是“a science dealing with the collection, analysis, interpretation and presentation of masses of numerical data”,即统计学是研究大量数字资料的收集、分析、解释和表达的一门学科。

根据我国学科分类标准(GB/T 13745—92),统计学和临床医学都是一级学科,分类代码分别为 910 和 320。目前,统计学在医学科学研究中的运用已越来越广泛、深入。在医学研究报告和医学期刊中,许多科学研究结果都直接用统计数据 and 统计图表表达,均数($\bar{X}$)和标准差(S)、率(p)和标准误(S_p)以及随机分组、 P 值等统计专业术语随处可见。那些无对照组,无统计结果,无 P 值的临床疗效评价研究结果,几乎不可能在高等级的医学期刊上发表。因此,作为医学科学研究方法学的医学统计学(medical statistics)就是一门运用统计学原理和方法研究医学科研中有关数据的收集、整理和分析的应用科学。由于医学统计学涉及伦理学等特殊问题,故有别于其他统计学。

医学科学研究的对象主要是人体以及与人的健康有关的各种因素。生物医学现象的一个重要特点就是普遍存在着变异(variation)。所谓变异(个体差异),系指相同条件下同类个体之间某一方面发展的不平衡性,系偶然因素起作用的结果。例如同地区、同性别、同年龄的健康人,他们的身长、体重、血压、脉搏、体温、红细胞、白细胞等数值都会有所不同。又如在同样条件下,用同一种药物来治疗某病,有的病人被治愈,有的疗效不显著,有的可能无效甚至死亡。因此,当存在着变异时,同样条件下的重复实验结果不会完全相同,具有不确定性,医生和医学研究人员必须学会用统计学来揭示生命现象中的必然性。哪里有变异,哪里就需

要统计学。

统计学在医学研究中的运用主要有三个方面:

1. 以正确的方式收集数据,如实验设计、调查设计等。
2. 描述数据的统计特征,如数据化简、统计指标的选择与计算、统计结果的表达等。
3. 统计分析及得出正确结论,如根据概率分布,对实验和观察结果存在的差异和关联作出统计推断。

§ 1.2 统计学的几个基本概念

1.2.1 随机事件与样本空间

在自然界和人类社会活动中,存在着两类不同的现象:确定性现象与随机现象。确定性现象在一定条件下必然发生(或必然不发生),例如,自然状态下向上抛一石子必然下落,同性电荷必不相互吸引,等等。随机现象是指在相同条件下重复进行的一系列试验,而每次试验的可能结果不止一个,试验的所有可能结果能事先明确,但在每次试验之前无法预知确切的结果,即呈现出不确定性。例如,在相同条件下抛同一枚硬币,其结果可能是花这面朝上,也可能是币值一面朝上,在每次抛掷之前无法肯定抛掷的结果是什么。随机现象在个别试验中呈现出不确定性;在大量重复试验中,又具有统计规律性。随机现象中的试验称为随机试验(random experiment),简称试验(experiment)。试验(以 E 表示)是一个广泛的术语,通过随机试验研究和揭示随机现象的统计规律性是统计学的重要任务之一。

在随机试验中,对一次试验可能出现也可能不出现,而在大量重复试验中却具有某种规律性的事情,称为此随机试验的随机事件(random event),简称事件。

随机试验中的每一个可能出现的结果都是一个随机事件,它们是该试验的最简单的随机事件,特称这些简单的随机事件为基本事件。

例如,在上述抛硬币试验中,“出现花面朝上”“出现币值朝上”就是试验的基本事件;在一颗骰子的抛掷试验中,“出现 1 点”“出现 2 点”...“出现 6 点”就是基本事件。

一试验中,除基本事件以外还有其他随机事件。例如,在一颗骰子的抛掷试验中,“出现偶数点”也是一个随机事件,它是由“出现 2 点”“出现 4 点”“出现 6 点”这三个基本事件所组成的,当且仅当这三个基本事件中有一个发生,“出现偶数点”这一事件才发生。

在试验 E 中必然会发生的事情叫做必然事件;不可能发生的事情叫做不可能

事件。例如掷一颗骰子,“出现点数不大于 6”是必然事件;“出现点数大于 6”是不可能事件。这两类事件具有确定性,不是随机事件,但可视为随机事件的特例。

为了便于研究随机试验 E , 将随机试验 E 的所有基本事件所组成的集合叫做样本空间(sample space), 记为 S 。例如掷一颗骰子试验, $S = \{1, 2, 3, 4, 5, 6\}$ 。每一个基本事件称为样本点。基本事件也就是集合 S 的元素。没有元素的集合称为空集, 记为 ϕ 。由于随机事件是基本事件, 或是由基本事件所组成的, 引入了样本空间 S 之后, 试验 E 的每一事件就是一个集合, 且是样本空间 S 的子集, 当且仅当子集中的一个样本点发生时, 试验 E 的事件即发生。例如上述的掷骰子试验中的事件 A : “出现偶数点”是由基本事件“2”“4”“6”所组成的, A 是 S 的子集, 即 $A = \{2, 4, 6\}$ 。必然事件就是样本空间 S ; 不可能事件就是空集 ϕ 。

1.2.2 频率与概率

设随机事件 A 在 n 次试验中出现 n_A 次, 比值

$$f_n(A) = \frac{n_A}{n} \quad (1.1)$$

称为事件 A 在这 n 次试验中出现的频率(frequency)。例如, 某病患者 40 名, 用某疗法治疗后, 其中 36 人痊愈, 治愈者占治疗人数的 $36/40 = 0.90$, 这是频率。频率能一定程度地反映事件 A 发生的可能性的, 但有随机波动性。当试验次数 n 逐渐增多时, $f_n(A)$ 在一个常数 $P(A)$ 附近摆动, 摆动的幅度随着 n 的增大将愈来愈小, 而逐渐稳定下来, 即当 n 很大时就有 $f_n(A) \approx P(A)$ 。这个常数 $P(A)$ 是客观存在的, 可用稳定值 $P(A)$ 来度量事件 A 发生的可能性的, 这就是定义事件的概率的客观基础。“频率稳定性”揭示了隐藏在随机现象中的规律性。这种规律性就是通常所说的统计规律性。

由上述可知, 当试验次数 n 逐渐增多, 频率 $f_n(A)$ 将稳定在一个常数附近, 这个常数称为事件 A 的概率(probability), 用 $P(A)$ 来表示。 $P(A)$ 反映了事件 A 发生的可能性的。例如, 上述的“抛硬币”试验的每一个基本事件的概率均为 $1/2$, “掷骰子”试验的每个基本事件的概率均为 $1/6$ 。设 E 是随机试验, S 是它的样本空间, 事件 A 的概率 $P(A)$ 必具有下列性质: ①对于每一事件 A 有 $0 \leq P(A) \leq 1$; ② $P(S) = 1$; ③对于两两互不相容(即两两不能同时发生)的事件 $A_k (k = 1, 2, \dots)$ 有

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (1.2)$$

随机事件的概率介于 0 与 1 之间, 即 $0 \leq P \leq 1$, 常用小数或百分数表示。 P 越接近 1, 表示事件发生的可能性越大, P 越接近 0, 表示事件发生的可能性越小。

$P=1$ 表示事件必然发生(必然事件); $P=0$ 表示事件不可能发生(不可能事件)。统计分析中的很多结论都基于一定可信程度下的概率推断,习惯上将 $P \leq 0.05$ 称为小概率事件,表示在一次实验或观察中该事件发生的可能性很小,可视为很可能不发生。

设 A 与 B 为随机试验 E 的两个事件,且 $P(A) > 0$,则称

$$P(B/A) = \frac{P(AB)}{P(A)} \quad (1.3)$$

为事件 A 发生的条件下事件 B 发生的条件概率。式(1.3)中的 $P(AB)$ 为事件 A 与 B 同时发生的概率。

1.2.3 随机变量与资料

设 E 是随机试验, $S = \{e\}$ 是它的样本空间,如果对于每一个 $e \in S$ 有一个实数 $X(e)$ 和它对应,这样就得到一个定义在 S 上的实值单值函数 $X(e)$,称 $X(e)$ 为随机变量(random variable),记 $X = X(e)$ 。 X 是随着试验结果不同而变化的一个量,因此,引入随机变量后,随机事件就可以通过随机变量来表示。

随机变量是建立在随机事件基础上的一个概念,随机变量随着试验的结果不同而取不同的值,在试验之前只知道它可能取值的范围,而不能预知它取什么值。由于试验的各个结果的出现(即随机事件的发生)有一定的概率,那么随机变量 X 也以一定的概率取各种可能的实数值。实践中常用的随机变量按其取值情况可分为两类,一类是离散型随机变量(discrete random variable),即只可能取有限个或无限可列个实数值的随机变量,如用某种药物治疗某种非传染性疾病,其疗效分为有效或无效,则在 n 个被治疗的患者中可能出现的有效例数 X 就是离散型随机变量,这里 $X = 0, 1, 2, \dots, n$; 另一类是连续型随机变量(continuous random variable),即可以在整个实数轴上取值,或至少有一部分值取自实数轴上的某个连续区间的随机变量,如某地正常成年男性的身高 X 即为连续型随机变量,其取值的实数区间可以是 $150.0\text{cm} \leq X \leq 195.0\text{cm}$ 。

观察单位(observed unit),是统计研究中根据需要确定的采集数据的最基本单位。它可以是一个人,一头动物,也可以是特指的一群人(例如一个家庭,一个幼儿园,一个自然村等);可以是一个器官,甚至一个细胞,一个采样点等。随机变量的取值称为变量值(value of variable)或观察值(observed value),由变量值构成资料(data)。资料分为以下几种类型:

1. 计量资料:计量资料(measurement data)又称定量资料(quantitative data)或数值变量(numerical variable)资料,为测定每个观察单位某项指标的大小,而获得的资料。测定结果表现为数值大小,一般有度量衡单位。为便于理论研究,其