普通高等教育“十二五”规划教材

郭秀花 主编

医学统计学与SPSS软件实现方法

Medical Statistics and SPSS Software Application

科学出版社

普通高等教育“十二五”规划教材

医学统计学与SPSS软件实现方法

郭秀花　主编

科学出版社

北京

内 容 简 介

医学统计学是我国高校医学专业各层次的必修课，是进行医学科学研究的一门方法学课程。目前我国高校缺少将医学统计学方法与常用的 SPSS 18.0 统计软件操作结合起来的教材。为弥补这一缺憾，特组织全国 12 所高校共同编写本教材。其主要特色有：第一，在内容安排上注重与医学科研实际相结合，注意统计知识的整体性与前后连贯性，将科研统计设计、数据管理与质量控制、数据统计分析几个步骤进行有机结合。第二，教材整体注重应用医学统计学基本理论与方法如何解决实际问题。重点在于什么样的问题采用怎么样的统计设计？什么样的实际数据，采用何种统计分析方法？如何对统计分析结果进行合理的解释？第三，结合 SPSS 18.0 统计软件窗口式操作简单、方便的特点，为学习者节省了大量的统计计算量和时间。第四，注重统计学方法的适用性与通用性，并将之与现代医学统计学理论相结合，为学习者开拓思维、处理高维及多因素统计分析模型的学习奠定了基础。第五，本教材的附录部分除了一般医学统计学书中给出的统计用表、关键词语的中英文对照外，还给出了各章练习题答案以及综合测试题，为课堂教学和自学提供了方便。

本教材可供临床、口腔、护理、检验、药学、中医学等专业本科生开设 30~60 课时的医学统计学教学使用，也可供各专业研究生开设基本（初级或中级）医学统计学课程教学使用。

图书在版编目(CIP)数据

医学统计学与 SPSS 软件实现方法/郭秀花主编. —北京：科学出版社，2012.8
普通高等教育“十二五”规划教材
ISBN 978-7-03-035152-4

I. ①医… II. ①郭… III. ①医学统计－统计分析－软件包－高等学校－教材 IV. ①R195.1-39

中国版本图书馆 CIP 数据核字(2012)第 162091 号

责任编辑：潘志坚 闵 捷 雷 旸 / 责任校对：张凤琴
责任印制：刘 学 / 封面设计：殷 靓

科 学 出 版 社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
江苏省句容市排印厂印刷
科学出版社发行 各地新华书店经销
*
2012 年 8 月第 一 版 开本：889 × 1194 1/16
2012 年 8 月第一次印刷 印张：19 3/4
字数：678 000
定价：38.00 元
(如有印装质量问题，我社负责调换)

《医学统计学与SPSS软件实现方法》编委会名单

主　编　郭秀花　首都医科大学公共卫生与家庭医学学院

副主编　刘美娜　哈尔滨医科大学公共卫生学院

黄水平　徐州医学院公共卫生学系

赵若望　包头医学院公共卫生学院

潘发明　安徽医科大学公共卫生学院

编　者　(按在书中出现的先后顺序)

郭秀花（首都医科大学）

刘美娜（哈尔滨医科大学）

罗艳侠（首都医科大学）

刘　芬（首都医科大学）

黄水平（徐州医学院）

曾　平（徐州医学院）

潘发明（安徽医科大学）

范引光（安徽医科大学）

肖焕波（首都医科大学燕京医学院）

杨兴华（首都医科大学）

张秋菊（哈尔滨医科大学）

郭淑霞（石河子大学医学院）

王　虹（首都医科大学）

李述刚（石河子大学医学院）

赵若望（包头医学院）

郝金奇（包头医学院）

尹素凤（河北联合大学）

贺　佳（第二军医大学）

李　霞（首都医科大学）

贾　红（泸州医学院）

孙　忠（天津医科大学）

王　媛（天津医科大学）

闫宇翔（首都医科大学）

吴立娟（首都医科大学）

刘启贵（大连医科大学）

胡冬梅（大连医科大学）

颜素容（北京中医药大学）

高　琦（首都医科大学）

学术秘书　闫宇翔

前言

医学统计学是统计学原理和方法与医学相结合的一门应用科学，是高校医学专业的必修课，是进行医学科学研究不可缺少的一门方法学课程。目前虽然有各种版本的医学统计学或统计软件操作手册方面的书籍，然而，医学统计学方面的教材偏重于理论，详细介绍原理、公式和计算过程，缺少直观而详细的统计软件操作方法；大多数医学统计学软件操作手册又偏重软件介绍，缺乏统计学基本概念与理论。我们编写的教材，是将医学统计学方法与常用的SPSS 18.0统计软件操作结合起来了，以弥补目前高校医学统计学教材的缺憾。

编写本教材是本着以下原则：第一，内容的科学性为主，兼顾理论的前瞻性。正确阐述医学统计学学科的科学理论和概念定义，在理论联系实际，以实例解释理论，对实践起到指导作用的基础上，注意将本领域的最新发展成果，把新技术、新方法纳入教材；第二，把握好写作条理性。注重教材的层次分明、条理清楚，教材体系能反映内容的内在联系及统计学的思维方式；第三，以本科学生或同等水平阅读能力群体为主要对象，并兼顾了课时数少的研究生医学统计学教学。从认知规律出发，富有启发性，便于学生学习，所选教学内容可以满足学生未来职业活动所需的最基本、最常用的理论知识和方法。第四，突出实践技能，强化应用。注重科研实际案例引入，统计方法计算以SPSS 18.0软件操作和结果解释为主，使学生真正掌握实践操作技能。

本教材共分为十六章：绪论、数据管理与SPSS统计软件简介、定量资料的统计描述、定量资料的参数估计与假设检验基础、定量资料的t检验、定量资料的方差分析、定量资料的非参数检验、定性资料的统计描述、定性资料的参数估计与χ^2检验、有序定性资料的统计分析方法、直线相关与回归、多重线性回归分析、Logistic回归分析、统计表与统计图、观察性研究设计、实验性研究设计。本教材后的附录部分有统计用表、各章练习题答案要点与综合测试题、英汉统计名词对照。本教材的作者在写作过程中，参考了大量的资料，涉及医学统计学、医学统计学习题集、SPSS软件操作方法等方面的书籍，与这些书籍相比，本教材具有以下特点：第一，在内容安排上注重与医学科研实际相结合，注意统计知识的整体性与前后连贯性，将科研统计设计（重点是统计设计）、数据管理与质量控制、数据统计分析几个步骤进行有机结合，强调数据管理与数据质量的必要性。第二，全书系统介绍医学统计学的基本概念、基本原理与基本方法，实用性强。重点在于，什么样的问题采用怎样的统计设计；什么样的实际数据采用怎样的统计分析方法；如何对统计分析结果进行合理的解释；等等。第三，结合SPSS 18.0统计软件窗口式操作简单、方便的特点，为学习者节省了大量的统计计算工作量和时间，从而将学习重点转移到对统计的三基的理解，而非数据公式的具体使用与计算。第四，注重统计学方法的适用性与通用性，并将之与现代统计学的理论相结合，如介绍有序列联表的统计学分析、多因素统计分析模型、非参数的多重比较等内容。第五，本书后面的附录部分除了一般医学统计学书中给出的统计用表、关键词语的中英文对照外，还给出了各章练习题答案以及综合测试题，为课堂教学和自学提供了方便。

Foreword

在本教材即将问世之际，我们首先感谢首都医科大学有关校领导、教务处各位领导对本教材编写工作的关心与指导；感谢全国高校素质教育教材研究编审委员会刘思祺主任、中国教师发展基金会教师出版专项基金办公室邱巍主任，对本教材出版工作给予的大力支持；感谢各高校同仁参加本书的编写；感谢首都医科大学公共卫生与家庭医学学院王嵬院长、郭爱民书记对出版本书提供的诸多指导；感谢学术秘书闫宇翔副教授为本书做了大量而繁杂的具体工作；同时，感谢我的研究生陶丽新、霍达、孙涛、潘蕾、周涛等同学对本书的所有例题进行了复核，并认真校对和排版书稿。最后，还要感谢我的丈夫和我可爱的女儿对我的理解和支持！

本教材供临床、护理、检验、药学等专业本科生，开设30~60课时的医学统计学教学使用；也可供各专业研究生开设60课时以内的医学统计学教学使用。虽然我们力求在编写内容、体例、实用等方面有新的创新与突破，但限于我们的学识和精力，本书的缺点在所难免，恳请广大读者批评指正（E-mail:guoxiuh@ccmu.edu.cn），以便再版时改正。

郭秀花

2012年1月于北京

目 录

Contents

目 录

Contents

目 录

Contents

第1章 绪 论

1.1 医学统计学概述

当今人类进入到信息化社会的21世纪，统计对我们每个人来说并不陌生，报纸杂志、电视广播、网络媒体等每时每刻都传递着很多统计数据和信息，我们也常听到很多关于“统计”的词汇。例如，据统计去年国民生产总值GDP增长率8.2%；某地人均寿命78.6岁；2月房屋销售量环比下降15.6%。有许多问题需要运用统计学给出答案。例如，治疗艾滋病的新药有效吗？明年中国股市涨跌趋势？体育彩票中奖的概率？子女是否像父母？目前居民对医疗改革政策的满意度是多少？可以说统计学的知识已经渗透到自然科学、社会科学以及人类生活的各个领域。在现代社会中，大到国家重大政策的制定，小到人们的日常生活，几乎都离不开统计学。

1.1.1 定义

1. 统计学的定义 在西方，统计学(statistics)一词，源出于state(国家、情况)，专指有关“国情”的学问，最初多用于文字记叙，后发展为数量比较，随着概率论思想与方法的引入，逐渐形成今天在理论与应用方面都已相当完备的独立学科。我国教育部1998年在《普通高等学校本科专业目录和专业介绍》中将统计学列为理学类一级学科。按照《教育部关于进行普通高等学校本科专业目录修订工作的通知》(教高〔2010〕11号)要求，《普通高等学校本科专业目录》修订工作2011年进入了公开征求意见阶段。

然而什么是统计？广义上是指通常人们所遇见的任何用数字、表格与图形所表达的一个事实，狭义上统计作为一门学科。什么是统计学(statistics)？统计学有其自身独有的知识体系和方法论。著名Webster国际大辞典中定义，统计学是“a science dealing with the collection, analysis, interpretation, and presentation of masses of numerical data”，即统计学是一门关于收集、分析、解释和表达数据的科学。

2. 医学统计学的定义 统计学与各个专业结合就形成数十个学科分支，如社会统计学、经济统计学、人口统计学、心理统计学、遗传统计学等。统计学的理论是随着人类社会生产的需要而产生，同时也随着人类社会生产的发展而更新，特别是近20年来，统计学的理论方法和应用方面得到迅速的发展，新的领域与统计学结合形成的新的分支如同雨后春笋般继续不断出现。如果把统计学应用在医学领域，即将统计学与医学相结合而形成的一个交叉科学，就形成了医学统计学。因此医学统计学的定义就是用统计学的原理和方法来研究医学领域中不确定性现象规律性的一门学科。目前，医学统计学是现在及未来一个世纪中最活跃、最有生命力的学科之一。

3. 医学统计学应用现况 现在，生物医学实验、临床试验、流行病学调查和公共卫生管理都要寻求统计学家的合作。医学科研基金申请要求有统计学家参与合作，申请书必须包含详尽的统计设计与分析；新药开发和报批必须依法执行统计学准则，递交统计分析报告；公共卫生项目的确立和验收，必须基于抽样调查的数据和完善的评价体系；医学杂志发布统计学指南，邀请统计学家审稿，严控论文的统计学缺陷。美国国立卫生研究院(National Institutes of Health，NIH)的基金申请明确要求基金合作者中有统计学家，并且在所立项中有统计学方面的内容。美国国家药品食品管理局(Food and Drug Administration，FDA)要求新药的研发试验中，必须有统计学家来指导研究的设计、数据的分析、报告的呈递等。总之，统计学思维和方法学已经渗透到医学研究和卫生决策之中。

但是，很多医学实际科研工作者对统计学的作用重视不够，突出表现在忽视医学科研设计、在统计分析时盲目套用统计分析方法、对统计分析结果解释时轻描淡写，一笔带过。把统计学当做无关紧

要的“修饰物”，严重影响了医学科研工作的科学性与严谨性。由于轻视或误用统计学而导致得出错误结论的例子并不鲜见。2002 年 11 月 9 日《科学时报》登载了军事医学科学院情报研究所胡良平教授提及的一个令人触目惊心的数据：全国各类医学期刊中，有统计学错误的论著竟占到 80%。2001 年西班牙的 Girona 大学的 Emili Garcia-Berthou 和 Carles Alcaraz 查阅了 Nature 上发表的 181 篇论文，发现 38%的文章至少有一处有统计学错误。2005 年 Nature Medicine 发表过一篇社论，题目为：“Statistically significant”，一开头就说“Nature 和 Nature Medicine 因为登载的某些文章统计分析欠佳而遭到公众批评”。

1.1.2 怎样学好医学统计学

许多学生习惯于传统的医学统计教学模式，往往是“填鸭灌输式”或“知识继承型”的教学方法，教师在上面讲，学生在下面听，忙于记笔记，死记硬背应付考试，以后科研中遇到统计学问题还是束手无策或误用滥用。在学习方法上我们提出如下建议：

1. 培养严谨、科学的态度 在医学科研中应用统计学的目的是要探究客观事物的规律性，提出或验证科学问题。有的人当应用统计处理实际资料得不到理想的结果(或阳性结果)时，就拼凑数据甚至修改数据，这是严重违背统计学主旨的，也是严重的学术造假行为。我们要遵从客观事实、认真分析原因，例如，各种因素是否考虑全面了？研究对象的选取是否合理？样本量是否足够大？指标选取的如何？收集资料的方法是否可靠？统计方法应用是否有误？统计计算是否正确？如果各个环节都没有问题，也许是我们最开始从专业上提出的科学问题就应该如此。学习医学统计学，就是要培养严谨、科学的态度。

2. 抓住三基，即基本概念，基本原理，基本方法 对复杂的公式的推导及公式的本身只需要了解一下其作用，而不必死记硬背其具体的形式，也不必深究其数学原理。在医学科学研究中所应用的统计学知识中约 70%是最基本的概念和经典的统计方法，其余则是较为复杂的、近代发展起来的统计理论和技术，而出现错误最多的却偏偏是前一部分。

3. 重视统计应用，把实际问题转化为统计问题 学习统计时一定要结合实例，最好从问题的原形入手，将其转化成统计问题，这是正确使用统计学的关键一步。然后根据设计类型、资料性质和分析目的，选择合适的统计分析方法进行资料处理。要经过从理论到实践、再从实践到理论的反复过程，循序渐进，才能逐渐掌握统计学，从而在运用统计学解决实际问题时，才能得心应手。能否把各种实际问题转化为统计问题，能否合理选用统计学方法、正确运用统计学的理论和方法解决实际问题，是学好医学统计学的难点所在，也是衡量医学统计学教学质量的“金标准”。

4. 熟练掌握统计软件的使用 目前可以用来进行数据分析的统计软件很多，如 SAS、SPSS、Stata、R 语言等。在解决实际问题时，要重视各种检验方法适用的前提条件及应用场合，可以忽略其具体的计算推导过程；要熟练地掌握一种统计软件(如最简单、直观的操作软件 SPSS)，学会正确使用统计软件和正确选择统计方法，对软件输出结果及统计学结果学会正确解释。随着现代统计学和计算机技术的迅猛发展，一些新的统计学方法和技术逐渐成熟并得到广泛应用，统计软件的功能也日益强大，并促使医学研究向深度和广度发展。

1.2 医学统计工作的基本步骤

医学科学研究的全过程应有统计方法与统计工作相伴随，统计方法已广泛渗透到医学科研的各个环节，统计工作就是统计方法在医学科研中恰当和正确地使用。从统计学角度来说，统计工作的基本步骤是对科研项目进行设计、收集资料、整理资料和分析资料。

1.2.1 设计

设计(design)是在医学科学研究时，根据其研究的问题的目的，确定计划及方案的过程，是科学研究工作的纲领和完成研究工作的关键环节。

设计包括专业设计和统计设计。专业设计是从专业角度考虑实验的科学安排，是科学研究的基础，包括选题，建立假说，确定研究对象和技术方法等；统计设计是在明确研究目的的前提下，从统计学角度对资料进行收集、整理和分析提出全面具体的计划和要求，作为统计工作实施的依据，用尽可能少的人力、物力和时间获得准确可靠的结论。对于实验性研究的统计设计，根据研究目的制定研究方案，包括研究对象的纳入标准和排除标准、样本获取方法、实验与对照的分组、确定观察指标、实验过程中的质量控制和拟使用的统计方法等。对于观察性研究的统计设计，采用调查问卷或访谈的方法，直接从某社会群体中收集资料，通过对资料的统计分析研究科研问题的规律。

无论实验性研究还是调查性研究的统计设计，都强调了如何获得符合研究目的的可靠研究资料，正确的整理资料过程和分析方法，使结果能很好地回答所研究的问题。具体内容应体现为：明确同质的研究对象；明确取得原始资料的方法；如何整理资料的过程；计算哪些指标；用何种统计推断方法及对结果的预测。例如，研究补钙对绝经期妇女骨密度的影响：研究对象为绝经1年以上、年龄在50~65岁之间，排除影响骨密度的相关疾病、手术史、服用过激素类药物等因素的绝经期妇女；通过调查表和干预实验获得原始资料；利用计算机建立数据库进行资料的整理；计算指标的均数、标准差和率等；统计推断方法主要是方差分析、χ^2检验和多因素回归分析等；结果预测为排除干扰因素后给予不同剂量钙绝经期妇女的骨密度有差别。

1.2.2 收集资料

收集资料(data collection)是获得研究所需原始数据的过程，要根据研究目的与设计确定。实验性研究收集资料主要是通过专项实验，如动物实验，临床观察实验；调查性研究收集资料主要是通过专题调查。无论何种途径收集到的资料，都应强调它的准确性、完整性。医学科学研究原始资料的来源可以有：

1. 报表资料 医疗卫生领域里的各种报表，如传染病报表、疾病监测报表、医院年度统计报表、卫生统计年鉴等。研究中国传染病的疾病负担，要收集几年内中国疾病预防控制中心或卫生部传染病的报告数据；对十年后中国卫生技术人员中医生和护士人数进行预测，要对近20年的卫生统计年鉴内中国卫生技术人员中医生和护士人数进行收集，建立数据集进行预测。

2. 医疗、卫生机构的日常工作记录 如住院病例、经常性工作记录和数据库等。疾病治疗质量评价的研究中，确定了评价指标后，要对医院住院患者的病例数据进行收集，利用统计分析方法进行影响因素调整和治疗质量评价。

3. 专题研究的实验数据和调查资料 如补钙对280名绝经期妇女骨密度影响的数据收集有两个部分：一是通过调查表调查绝经期妇女的一般情况、饮食情况、体育锻炼情况、生育史、心理健康与应对等调查资料；二是实验研究数据，实验分为4组，每组70人，信息干预组(只透露本人检查结果，不给干预措施)，其余三组分别给A、B、C三种含不同剂量钙的奶粉，补钙一年后和两年后分别测其骨密度值，血液、尿液中的实验室指标。

1.2.3 整理资料

整理资料(data sorting)是指对收集到的原始资料进行归类整理汇总的过程。即有目的地对收集到的原始资料进行科学加工，使资料系统化、条理化，以便进行统计分析。包括三方面的内容：

1. 数据净化(data cleaning) 对数据进行去伪存真的过程，即对原始数据进行检查、核对、纠错和改正。

2. 逻辑检查(logical check) 通过计算机对数据进行检查与核对的过程。根据逻辑关系、常识和专业背景知识，对所研究的资料进行检查与核对。对产生怀疑的数据，要进行深入核查予以纠正。

3. 统计核查(statistical check) 为了进行统计分析，需要对原始数据进行加工，将其转化为频数分布表(frequency distribution)数据，可以根据数据间的关联性和频数分布图或表等进行核查。

整理资料主要步骤是审核资料，拟整理表和归纳汇总。在补钙对绝经期妇女骨密度的影响研究中，根据补钙前和补钙一年和两年后的一般情况、饮食情况、体育锻炼情况，骨密度值及实验室等指标。

建立Epidata数据库，采用双向比对的形式对数据进行录入，检查与核对。将数据导入SPSS分析软件内，再进一步对数据进行逻辑检查。对数据进行加工、整理、归纳汇总等。

1.2.4 分析资料

分析资料(data analysis)就是对整理的资料进行统计分析，获取资料中有关信息的过程。包括统计描述(statistical description)和统计推断(statistical inference)两个方面。统计描述是通过计算有关的统计指标，对资料进行全面概括地描述，即统计指标的计算和统计图表的绘制。统计推断是从样本中的信息推断总体特征，包括两部分：一是参数估计(estimation of parameter)，用样本统计量估计总体参数；二是假设检验(hypothesis test)，用样本信息检验关于总体之间的差别。

补钙对绝经期妇女骨密度影响的分析资料中：统计描述为计算骨密度的均值和标准差，根据腰椎骨密度的T<–2.5为骨质疏松计算患病率；绘制统计表和统计图。对调查数据的统计推断是以骨密度值为主要指标，用多重线性回归分析一般情况、饮食因素及其他因素与骨密度之间的关系；将人群分为骨质疏松症和非骨质疏松症两组，用Logistic回归分析骨质疏松症的相关危险因素。对实验干预数据的统计推断，按照重复测量设计，用方差分析研究三次骨密度值与不同干预分组及与时间变化的关系。

1.3 医学统计学中的几组基本概念

1.3.1 总体和样本

总体(population)是根据研究目的确定的，同质个体所构成的全体。总体分为有限总体(finite population)和无限总体(infinite population)：总体的个体可数，研究单位是有限的，可以确定为有限总体。总体的个体不可数，研究单位是无限的，没有时间、空间限定，研究单位的全体只是理论上存在为无限总体。如补钙对绝经期妇女骨密度的影响，全部绝经期妇女是研究对象，这是有限总体。研究松花江水中甲基汞含量，全部松花江水是研究对象，这是无限总体。无论是有限总体还是无限总体，在医学科学研究的过程中都不可能做到将总体中的个体都进行研究。补钙对绝经期妇女骨密度的影响研究，要在绝经期妇女中随机抽取一部分作为研究对象，松花江的上中下游分别抽取一部分水样作为研究对象，然后根据研究结果推断绝经期妇女的骨密度和松花江水中甲基汞含量。

样本(sample)是从总体中随机抽取的部分个体，一定要具备代表性和可靠性。统计学把描述总体特征的指标称为参数(parameter)，描述样本特征的指标称为统计量(statistic)。医学科学研究的目的是由样本推断总体，依据统计量的特征或性质对总体参数作出结论，阐明总体的特征与规律。

代表性(representation)就是要求样本能够充分反映总体的特征。根据研究目的对总体有一个明确的规定后，样本必须是从总体中随机抽取出来，抽取样本的过程称为抽样(sampling)。随机(random)即需要保证总体中的每个个体有相同的可能被抽出来作为样本，要避免主观的偏性。必须指出的是，随机化抽样绝不等于随意抽样。为了保证抽样的随机性，可用抽签法、机械抽样法、分层抽样法、随机数字表及计算机随机化抽样等方法。

可靠性(reliability)就是样本一定属于所规定的总体范围内，样本数量要足够大。样本包含的个体数目称为样本含量(sample size)。由于个体之间存在差异，只有观察的样本含量达到一定数量才能体现出其客观规律性，当样本的例数过少时，得到的结论是不可靠的。样本的含量越大，结论可靠性会越大，但随着例数增加，需要的人力和物力也相应增加，所以应以“足够”为准。究竟需要多少例数，与所观察的指标的变异程度有关，需要参考相应的教材对样本含量进行计算。

1.3.2 同质与变异

数据是总体或样本中每个个体某特征的全部取值。例如，280名绝经期妇女的骨密度数据，100名同性别、同年龄的小学生的身高(或体重)数据。这些数据处于绝经期妇女和同性别、同年龄小学生的同一总体中，具有同质性(homogeneity)。然而，每个个体的骨密度、身高(或体重)间又存在差异，这种现象称为变异(variation)。医学研究的对象是有机的生命体，其机能十分复杂，不同的个体在相

同的条件下，对外界环境因素的影响可以发生不同反应。例如，给绝经期妇女补相同剂量的钙，测其骨密度值可能会各不相同；在相同的条件下测同年龄、同性别的健康人的脉搏、呼吸、体温等生理指标可以有很大差异；在临床治疗中，用同样的药物治疗病情相同的病人，疗效也不尽相同；在实验室里，同种动物之间也会有差异。

总体是由同质的个体构成，个体之间没有变异就无需统计分析。统计学的任务就是在变异的基础上描述同一总体的同质性，研究不同总体的异质性。

1.3.3 变量与资料

同质研究对象的某特征值具有变异性，构成了研究的变量(variable)，变量全部或部分的测量值构成资料(data)。在补钙对绝经期妇女骨密度的影响研究中，可以得到骨密度值、血液、尿液中指标等变量，280 名绝经期妇女的这些变量测量值构成资料；两家医院对肺结核治疗结果，可以得到病人的年龄、性别、体重和痰涂片等变量，两家各有 35 名患者的这些变量测量值构成资料。

统计分析中识别变量和资料的类型非常重要，决定了统计分析方法的选择，换句话说不同类型的资料要用不同的统计方法去分析。资料类型分为计量资料(measurement data)、计数资料(enumeration data)和等级资料(ranked data)；变量分为定量(quantitative)变量和定性(qualitative)变量。

1. 计量资料 可用定量方法获得变量的测量值，具有计量单位，一般是由定量变量构成。定量变量分离散型变量(discreate variable)和连续型变量(continuous variable)。离散型变量只能取整数值，如一年中的手术病人数，新生儿数；连续型变量可以取实数轴上的任何数值，“连续”是指该变量可以在实数轴上连续变动，如年龄、身高、体重、骨密度等。

2. 计数资料 按属性或类别分组获得变量的个数，由定性变量中的分类变量构成。定性变量有分类变量(categorical variable)和有序分类变量(ordinal categorical variable)。最常用的分类变量如性别的男女、疾病的有无、结局的生死等是分类变量中的二分类变量(binary variable)；职业分工、农、商、学、兵，血型分 O 型、A 型、B 型、AB 型等是分类变量中的多分类变量(polytomous variable)。

3. 等级资料 按程度或级别分组获得变量的个数。如结核患者的痰涂片结果为阴性、可疑、阳性，临床体检或实验室检验常用–、±、+、++和+++来表示测量结果，这些均为有序分类变量构成等级资料。

为了研究需要或数据分析方便，有时要对资料进行转换。一般是将计量资料转为计数资料或等级资料。例如，血红蛋白水平为计量资料，根据贫血判断标准，孕妇血红蛋白水平小于 110g/L 为异常，可将孕妇分为正常与贫血两类，构成计数资料。按 110g/L、70g/L、40g/L 划分，可把血红蛋白水平转换为正常、轻中度贫血、重度贫血、极重度贫血四类，构成等级资料。

1.3.4 误差与偏差

误差(error)是指观测值与真实值之差，以及样本统计量与总体参数之差，在没有真实值(总体参数)的情况下误差既是偏差。误差主要有以下三类。

1. 过失误差 由于科研工作者的失误或过错造成的误差。这是科研工作中绝对不允许出现的误差，科研工作者对科研工作要有严谨的工作态度，以达到消除过失误差的目的。

2. 系统误差 在收集资料过程中，由于仪器未经校正、医生掌握疗效标准偏高或偏低等原因，可造成观察结果的偏大或偏小，称作系统误差(systematic error)。系统误差影响原始资料的准确性，其原因是可以找到的，在医学科研中必须控制。统计设计中随机分组、设立对照等是控制系统误差的重要手段。例如，在补钙对绝经期妇女骨密度的影响研究中，将 280 名绝经期妇女随机分为四组，设立一个对照组，其余三组给不同剂量钙的奶粉。

3. 随机误差 在没有过失误差和系统误差情况下，仍存在的误差即随机误差。随机误差有很多的表现形式，在收集原始资料过程中，即使仪器初始状态及标准试剂已经校正，但由于各种偶然因素的影响仍会造成同一对象多次测定的结果不完全一致，这种误差往往没有固定的倾向，称为随机测量误差(random measurement error)。对于这种误差应采取措施，尽最大可能控制，至少应控制在一定的