卫生部规划教材

高等医药院校教材
供预防医学类专业用

卫生统计学

第三版

杨树勤 主编

人民卫生出版社

高等医药院校教材

（供预防医学类专业用）

卫生统计学

第三版

杨树勤　主编

（按章节顺序）

杨树勤（华西医科大学）

周有尚（同济医科大学）

倪旱雨（华西医科大学）

杨瑞璋（哈尔滨医科大学）　编写

杜养志（中山医科大学）

何大卫（山西医学院）

王绍贤（北京医科大学）

顾杏元（上海医科大学）

人民卫生出版社

编 写 说 明

本书是根据全国预防医学专业教材评审委员会对第三轮教材的编写要求，在总结使用本书第二版经验的基础上改写的，主要有以下特点：

1．从培养目标出发，精选“三基”内容。①尽量避免与相关学科（特别是流行病学、儿少卫生学和社会医学）的内容重复；将原22章改为16章，缩减了字数。②力求反映学科发展，更新了部分概念和方法。③删除了部分数学推理，增补了若干应用空缺。如删除了多元回归中的矩阵讲述；加强了计算机应用，补充了多个样本均数(率)比较的样本例数估计等。④使讲述更加严谨，理论更加结合实用，文字更加简明易读。

2．突出能力培养。①加强统计思维方法和能力培养。多由实例入手引出概念和推理，培养学生举一反三的能力；多用流程图、示意图、综合归纳表及章末小结等形式，帮助学生抓住要点，了解统计过程和方法的内在联系，提高自学能力。②加强统计分析和解决问题的能力培养。各种统计方法的用途、条件、方法、步骤明确，有实例说明；完善了习题，有助于培养综合分析能力。

本书共16章：第1章绪论，第2～13章基本统计方法，第14～16章健康统计，是按100学时设计的，有“*”节、段，可供选学。本书还提出了一些可进一步学习的问题，指出参考书目，如写明参见〔1:3～13〕，表明此问题可查阅书末所附参考书1，第3～13页。还有统计用表、常用统计软件包简介、实习题和英汉统计学词汇四个附录。

本书改编除吸收了第一、二版的编写经验外，还引用了前版的部分资料，谨向第一、二版中未参加第三版的作者致谢。

本书编写过程中，承华西医科大学公共卫生学院院长倪宗瓒教授给予支持；大连医学院胡克震教授，湖南医科大学黄正南教授，华西医科大学钱建明教授，重庆医科大学周燕荣教授，华北煤炭医学院郑俊池副教授等提出了宝贵意见；华西医科大学陈彬副教授和刘关键讲师任本书学术秘书，并负责书中计算的核对；刘关键讲师会同潘晓平讲师、黄志勇和陈健美老师在计算机上完成了本书稿的编辑、修改和打印；陈峰讲师协助了绘图和校对；谨致以衷心地感谢。我们还要感谢那些关心、支持和利用各种方式对本书第二版提出宝贵意见的读者，这些意见对本版改编有很大帮助。

这次改编，全体编者力图提高质量，但限于水平，一定还有不少缺点、错误，欢迎使用本书的师生和广大读者批评指正。

杨 树 勤

1992年10月1日于华西医科大学

目　录

此为试读,需要完整PDF请访问: www.ertongbook.com

第一章 绪 论

1.1 卫生统计学的定义和内容

统计学（statistics） 辩证唯物主义认为：世界是物质的，物质是运动的，运动是有规律的。为了能动地改造世界，首先要认识世界，研究物质世界的客观规律。需知，事物是在质和量的密切联系中发展的，研究问题既要注意其质的方面，也要注意其量的方面；事物的数量表现，既受本质规律的制约，又受许多偶然因素的影响，而且往往是偶然性（不确定性）掩盖了必然性，妨碍了我们对规律性的认识。例如正常儿童的身高总是随年龄增加而增加的，但也有一些年龄较大的儿童比年龄较小的儿童还矮。统计学是研究数据的搜集、整理与分析的科学，面对不确定性数据作出科学的推断，因而统计学是认识世界的重要手段。

当今信息社会，对有效地搜集数据，进行精确分析和可靠推断，作出科学决策，有着广泛的需求，这就使得统计学原理和方法几乎应用于科技所有领域，以及生产、生活和国民经济的各个部门，产生了许多应用性分支，如工业统计、农业统计和卫生统计等。

卫生统计学（health statistics） 它是以医学，特别是预防医学和卫生管理学的理论为指导，用统计学的原理和方法研究医学，侧重于预防医学和卫生事业管理中数据的搜集、整理与分析的一门应用科学。

卫生统计学的主要内容包括三个方面：① 卫生统计学的基本原理和方法。包括研究设计和数据处理中的统计理论和方法（详 1～13 章）。② 健康统计。包括医学人口统计、疾病统计（详 14～16 章）和生长发育统计（详儿少卫生学）等。③ 卫生服务统计。包括卫生资源、医疗卫生服务的需求和利用、医疗保健制度和管理等中的统计问题（详社会医学）。

电子计算机的发展和普及应用，为大量的信息储存与检索，复杂的数据处理，特别是多因素分析，以及抽样模拟等提供了条件，也促进了卫生统计学的发展。

1984 年起我国施行了“中华人民共和国统计法”，1992 年卫生部又发布了“全国卫生统计工作管理办法”，为有效地、科学地组织统计工作提供了法律保证。

目前全国各级卫生部门正在建立和健全卫生统计信息系统，全面发挥统计工作的信息、咨询和监督功能：统计信息功能指通过统计信息管理系统，采集、处理、传递、存储和提供统计信息；统计咨询功能指运用各种科学分析方法和现代计算手段，进一步开发统计信息资源，为科学决策和管理提供咨询建议与对策方案；统计监督功能指根据监测信息与反馈信息，及时、准确地对事物运行状态进行监督评价。

卫生医师在从事预防保健工作中，经常需要调查了解服务人群的健康状况，应保持与统计信息系统的密切联系（包括索取与提供有关统计信息），提出预防保健措施，并评价其效果，还必须根据工作需要开展科学研究。为此，预防医学专业学生必须学习卫生统计学。

1.2 统计工作的步骤

统计学是统计工作实践的经验总结，但它又对统计工作的全过程起指导作用，这个全过程可分为以下四个步骤：

1. **设计** (design) 设计之前，先要对研究的问题有较多的了解。为此，需要广泛查阅文献，了解实际情况，而且，常要与有关专家共同协作。设计的内容包括资料搜集、整理和分析全过程总的设想和安排。例如，什么是研究目的和假说？什么是观察对象和观察单位？需要搜集哪些原始资料？用什么方式和方法取得这些原始资料？怎样对取得的资料作进一步的整理汇总和计算统计指标？如何控制误差？预期会得到什么结果？需要多少经费等。凡此种种，都要结合实际，周密考虑，妥善安排。设计是后续步骤的依据，是最关键的一环。详见第十二章和第十三章。

2. **搜集资料** (collection of data) 任务是取得准确可靠的原始数据。卫生工作中的统计资料主要来自三个方面：① 统计报表。如法定传染病报表，职业病报表，医院工作报表等。这是国家规定的报表，由国家统一设计，要求有关医疗卫生机构定期逐级上报，提供居民健康状况和医疗卫生机构工作的主要数字，作为制定卫生工作计划与措施，检查与总结工作的依据。报表要做到完整、准确、及时。首先要保证基础资料的质量，如法定传染病报告卡是法定传染病报表的基础资料，要提高基层卫生人员的认识和责任感，要加强对漏报、重报和错报的检查。② 经常性工作记录。如经常性的卫生监测记录、健康检查记录等。要做到登记的完整、准确。病历是医疗工作的重要记录，分析时应注意其局限性（如不能反映一般人群特征）。③ 专题调查或实验。详第十二章和第十三章。

3. **整理资料** (sorting data) 任务是净化原始数据，使其系统化、条理化，便于进一步计算指标和分析。首先是资料清理 (data cleaning)。因为无论是调查或实验的原始记录和计算机录入过程，常会有错误，必须经过反复地检查和核对，这是需要耐心从事的基础工作，特别是数据较多时，一定要在修正错误，去伪存真之后，再开始按分析要求，分组汇总资料。计算机汇总逐渐多用，但数据规模较小时也可采用手工汇总。此项工作将在 12.3 节中作进一步讲述。

4. **分析资料** (analysis of data) 目的是计算有关指标，反映数据的综合特征(亦称综合指标)，阐明事物的内在联系和规律。统计分析包括：① 统计描述 (descriptive statistics)。指用统计指标、统计表、统计图等方法，对资料的数量特征及其分布规律进行测定和描述，不涉及由样本推论总体问题。② 统计推断 (inferential statistics)。指如何抽样，以及如何由样本信息推断总体特征问题。详见第二章至第十一章。

以上四个步骤是紧密联系，不可分割的整体，任何一步的缺陷，都会影响统计分析的结果。

1.3 统计中的几个基本概念

1. **变量** (variable) 无论用何种方式搜集资料，都要先确定观察单位 (observation unit)，亦称个体 (individual)，它可以是一个人、一个家庭、一个地区、一个样品、一个采样点等；然后对每个观察单位的某项特征进行测量和观察，这种被观察单位的特

征称为变量（习惯上亦称指标，但勿与前述综合指标相混）。例如，以人为单位，调查某地某年新生儿，性别变量的观察结果有男有女；生母年龄变量的观察结果，母龄有大有小；又如给同种属、同性别、年龄相近的小白鼠，喂以同种饲料，过一定时间，观察每鼠所增体重变量，结果各鼠增重不等。这种个体间的差异，通称变异（variation）。变异是宇宙事物的个性反映，在生物学和医学现象中尤为重要，它使统计学有特殊用武之地。个体差异，来源于一些未加控制或无法控制，甚至不明因素所致的随机误差。以上可见，变量的观察结果可以是定量的，也可是定性的，通称为变量值（value of variable）或观察值（observed value, observation）。按变量值是定量的还是定性的，可将变量分为以下类型。不同类型的变量应采用不同的统计分析方法。

（1）数值变量（numerical variable）或称定量变量，其变量值是定量的，表现为数值大小，一般有度量衡单位，亦称计量资料。如调查某地某年7岁女童的身体发育状况，以人为观察单位，每个人的身高（cm）、体重（kg）、血压（kPa）、坐高指数（%，坐高/身高）等均属数值变量。常用第二章至第五章和第十章的统计分析方法，有时亦用第九章的方法。

（2）分类变量（categorical variable）或称定性变量，其变量值是定性的，表现为互不相容的类别或属性，有两种情况：

1）无序分类（unordered categories）。包括：① 二项分类。如检查某小学学生大便中的蛔虫卵，以每个学生为观察单位，结果可以是蛔虫卵阳性或阴性；又如观察用某药治疗某病患者的治疗结果，以每个患者为观察单位，结果分为治愈与未愈两类。两类间互相对立。② 多项分类。如观察某人群的血型，以人为观察单位，结果分为A型、B型、AB型与O型，为互不相容的多个类别。无序分类变量的分析，应先分类汇总，计观察单位数，编制分类资料的频数表，亦称计数资料。常用第六章至第八章的统计分析方法。

2）有序分类（ordinal categories）。各类之间有程度的差别，给人以"半定量"的概念，亦称等级资料。如测定某人群血清反应，以人为观察单位，结果可分－、±、＋、＋＋四级；又如观察用某药治疗某病患者的治疗结果，以每个患者为观察单位，结果分为治愈、显效、好转、无效四级。有序分类变量的分析，应先按等级顺序，分类汇总，计观察单位数，编制等级资料的频数表。常用第九章、8.3节和10.5节的统计分析方法。

根据分析需要，各类变量可以互相转化。如以人为观察单位观察某人群成年男子的血红蛋白量（g/L），属数值变量；若按血红蛋白正常与异常分为两类，可按二项分类变量处理；若按血红蛋白量的多少分为五个等级：重度贫血、中度贫血、轻度贫血、正常、血红蛋白增高，可按等级变量处理。有时亦可将分类变量数量化，如将分多项的治疗结果转化为评分，分别用0、1、2…等表示，则可按数值变量处理。

2. **总体**（population）**与样本**（sample）总体是根据研究目的确定的同质观察单位的全体，更确切地说，是同质的所有观察单位某种变量值的集合。例如调查某地1992年正常成年男子的红细胞数，则观察对象是该地1992年的正常成年男子，观察单位应是每个人，变量是红细胞数，变量值是每人测得的红细胞数，该地1992年全部正常成年男子的红细胞数就构成一个总体。它的同质基础是同一地区、同一年份、同为正常成人、同

为男性。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，如研究贫血患者用某药治疗后的疗效，这里总体的同质基础是同为贫血患者，同用某药治疗，总体包括设想用该药治疗的所有贫血患者的治疗结果，是没有时间和空间范围限制的，因而观察单位数无限，称为无限总体（infinite population）。

医学研究中，很多是无限总体，要直接研究总体的情况是不可能的。即使对有限总体来说，若包含的观察单位过多，也要花费很大的人力、财力，有时也是不必要的和不可能的。如罐头食品的卫生检查，不可能将所有生产的罐头一一加以检验。所以在实际工作中经常是从总体中抽取样本，目的是用样本信息来推断总体特征。样本是从总体中随机抽取部分观察单位，其实测值的集合。如上例，可从某地 1992 年的正常成年男子中，随机抽取 144 人，分别测得其红细胞数，组成样本。所谓随机抽样，就是按随机的原则获取样本，避免研究者有意或无意地给样本带来偏性。有多种随机抽样方法可供选用，详见第十二章。样本包含的观察单位数称为样本含量或样本大小（sample size），医学上常称为样本例数。

3. **概率**（probability） 医学研究的现象，绝大多数是随机现象。例如用相同治疗方法治疗某病患者，只知道治疗转归可能为治愈、好转、无效、死亡四种结果，但对一个刚入院的该病患者，治疗后究竟发生哪一种结果是不确定的。这里的每一种可能结果都是一个随机事件，亦称偶然事件，简称事件。概率是描述随机事件发生的可能性大小的数值，常用 P 表示。比如本例将结果为“治愈”这个事件记为 A，则该患者治愈的概率可记为 $P(A)$，或简记为 P，这是一个很有意义的，医生颇为关心的未知数值。假如我们用 200 例的样本，求得治愈率为 75%，这只是一个频率。在实际工作中，当概率不易求得时，只要观察单位数充分多，可以将频率作为概率的估计值。但在观察单位数较少时，频率的波动性是很大的，用于估计概率是不可靠的。

随机事件的概率在 0 与 1 之间，即 $0 \leqslant P \leqslant 1$，常用小数或百分数表示。$P$ 越接近 1，表明某事件发生的可能性越大，P 越接近 0，表示某事件发生的可能性越小。严格说，$P=1$，表示事件必然发生，$P=0$，表示事件不可能发生，它们是确定性的，不是随机事件，但可把它们看成随机事件的特例。统计上的很多结论都是带有概率性的。习惯上将 $P \leqslant 0.05$，或 $P \leqslant 0.01$，称为小概率事件，表示某事件发生的可能性很小。关于概率知识的进一步了解参见〔1:3～13〕，或其他概率论与数理统计书。

1.4 学习卫生统计学应注意的问题

本课的教学目的是为学生在校学习其他学科，毕业后从事预防医学工作和科研，打下必要的卫生统计学基础。为此，学习本课时，应注意：

1. 掌握卫生统计学的基本知识、基本概念、基本原理和基本方法，培养统计思维方法和能力。统计思维方法，即统计的逻辑思维方法。例如，由于事物存在个体差异，用样本推断总体就会出现误差，但这种误差是有规律性的，就产生了统计推断的理论；懂得了假设检验的逻辑推理，就能理解统计结论的概率性。只有在学习基本理论知识中，逐步培养统计思维能力，才能提高统计自学能力，才有利于进一步获取知识。

2. 掌握调查设计和实验设计的原则，培养搜集、整理、分析统计资料的系统工作能

此为试读，需要完整PDF请访问：www.ertongbook.com

力。首先要重视原始资料的完整性和准确性，对数据处理持严肃、认真、实事求是的科学态度，反对伪造和篡改统计数字；要正确应用统计方法，对统计公式，只要求了解其意义、用途和应用条件，不必深究其数学推导。

3. 掌握群体健康的评价方法，学会用医学人口统计和疾病统计等方面的统计指标，综合评价人群健康状况，为卫生决策提供统计信息。

总之，应多联系实际，结合专业，分析评价卫生工作、医学文献和医学科研中的统计问题，才能学好卫生统计学。

小　结

1. 卫生统计学是统计学的一个分支学科，是应用统计学的原理和方法，研究医学和卫生事业管理中数据搜集、整理与分析的一门应用科学。它的主要内容，包括卫生统计学的基本原理和方法、健康统计和卫生服务统计。

2. 统计工作可分为设计、搜集资料、整理资料和分析资料四个步骤，任何一步的缺陷都会影响分析结果，而设计是资料搜集、整理和分析全过程总的设想和安排，是最关键的一步。

3. 观察单位的研究特征称为变量（或指标），变量的观察结果称为变量值，不同的变量用不同的统计分析方法。根据分析需要，各类变量可互相转化。

表 1.1　变量的类型及其分析方法(详见本书有关章节)

变量类型	变量值表现	例	分析方法
数值变量	定量(数值的大小)	红细胞数(10^{12}/L)	2～5章、10章、9章
分类变量	定性(不相容的类别)		
无序分类：二项	对立的两类	疗效：治愈、未愈	6～8章
多项	不相容的多类	血型：A、B、AB、O	6章、8章
有序分类(等级)	类间有程度差别	疗效：治愈、显效、好转、无效	9章、8.3节、10.5节

4. 总体是同质的所有观察单位某种变量值的集合，样本是从总体中随机抽取部分观察单位，其实测值的集合。抽样的目的是用样本信息推断总体特征。

5. 概率P是描述随机事件发生的可能性大小的数值，$0 \leqslant P \leqslant 1$，用小数或百分数表示。习惯上将$P \leqslant 0.05$或$P \leqslant 0.01$作为小概率事件。

6. 学习卫生统计学重点是掌握本学科的基本知识、基本概念、基本原理和基本方法；培养统计思维能力和工作能力；培养重视原始资料的完整、准确，对数据处理持严肃认真的科学态度。

（杨树勤　编）

第二章　频数分布的集中趋势与离散趋势

本章、下一章和第十一章将分别讲述数值变量资料（计量资料）的统计描述。

2.1　数值变量资料的频数表

1. 频数表（frequency table）**的编制**　为了解数值变量的分布规律，当观察单位较多时，可通过资料整理，编制频数分布表，简称频数表。

例 2.1　某市 1982 年 110 名 7 岁男童的身高（cm）资料如下，试编制频数表。

112.4	117.2	122.7	123.0	113.0	110.8	118.2	108.2	118.9	118.1
123.5	118.3	120.3	116.2	114.7	119.7	114.8	119.6	113.2	120.0
119.7	116.8	119.8	122.5	119.7	120.7	114.3	122.0	117.0	122.5
119.8	122.9	128.0	121.5	126.1	117.7	124.1	129.3	121.8	112.7
120.2	120.8	126.6	120.0	130.5	120.0	121.5	114.3	124.1	117.2
124.4	116.4	119.0	117.1	114.9	129.1	118.4	113.2	116.0	120.4
112.3	114.9	124.4	112.2	125.2	116.3	125.8	121.0	115.4	121.2
117.9	120.1	118.4	122.8	120.1	112.4	118.5	113.0	120.8	114.8
123.8	119.1	122.8	120.7	117.4	126.2	122.1	125.2	118.0	120.7
116.3	125.1	120.5	114.3	123.1	122.4	110.3	119.3	125.0	111.5
116.8	125.6	123.2	119.5	120.5	127.1	120.6	132.5	116.3	130.8

频数表的编制方法如下：

（1）找出观察值中的最大值、最小值和极差。本例最大值为 132.5cm，最小值为 108.2cm，最大值与最小值之差，称为极差，即

$$132.5\text{cm}-108.2\text{cm}=24.3\text{cm}$$

（2）按极差大小决定“组段”数、组段和组距。频数表一般设 10～15 个组段，观察单位较少时组段数可相对少些，观察单位较多时，组段数可酌情多些。常用极差的1/10取整作组距（class interval），取整只是为了方便资料整理汇总。第一组段要包括最小观察值，最后一个组段要包括最大观察值。本例，极差的1/10为 2.4cm，取整为 2.0cm，共分 13 个组段。

各个组段应界限分明，便于汇总。每个组段的起点称“下限”（low limit），终点称“上限”（upper limit）。本例身高是数值变量，按连续性资料处理，因此组段之间也应是连续性的。为避免含混，便于汇总，各个组段从本组段的“下限”开始，不包括本组段的“上限”，如表 2.1，第（1）栏：“108～”组段，包括身高在 108 至未满 110 的观察值，余仿此。注意：最末一组段应同时写出其下限和上限。

（3）列表划记。决定组段界限后，列成表 2.1 形式，将原始数据采用划记法或计算机汇总，得到各个组段的观察单位数（频数），表中第（1）、（3）栏即所需的频数表。

2. 频数分布的两个特征　从频数表可以看到频数分布的两个重要特征：集中趋势

(central tendency) 和离散趋势 (tendency of dispersion)。如由表 2.1 可见 110 名 7 岁男孩的身高有高有矮，但有一定的分布规律：① 身高向中央部份（即中等身高）集中，以中等身高者居多，是为集中趋势；② 从中央部份到两侧（即由中等身高到较矮或较高）频数分布逐渐减少，是为离散趋势。集中趋势和离散趋势是频数分布的两个重要侧面，测定其集中趋势和离散趋势就可较全面地分析所研究的事物。

表 2.1 110名7岁男童身高的划记表

身高组段 (1)	划记 (2)	频数 (3)
108～	一	1
110～	下	3
112～	正下	9
114～	正下	9
116～	正正正	15
118～	正正正下	18
120～	正正正正一	21
122～	正正下	14
124～	正正	10
126～	下	4
128～	下	3
130～	丁	2
132～134	一	1
合计	一	110

3. 频数分布的类型 频数分布又可分为对称分布和偏态分布两种类型。所谓对称分布是指集中位置在正中，左右两侧频数分布大体对称，如表 2.1 所示。若将其绘制成图 2.1，则更为清楚。所谓偏态分布是指集中位置偏向一侧，频数分布不对称，如一些以儿童为主的传染病，患者的年龄分布，集中位置偏向年龄小的一侧，称为正偏态分布；又如一些慢性病患者的年龄分布，集中位置偏向年龄大的一侧，称为负偏态分布。本书将在 4.7 节中作进一步讲述，参见图 4.2。不同类型的分布，应采用相应的统计分析方法。

4. 频数表的用途

(1) 已如上述，频数表可揭示资料的分布特征和分布类型。因而在文献中常将频数表作为陈述资料的形式。

(2) 便于进一步计算指标和统计分析处理，详下文。

(3) 便于发现某些特大或特小的可疑值。例如有时在频数表的两端，出现连续几个组段的频数为 0 后，又出现一些特大值或特小值，使人怀疑这些数值是否准确，需要进一步检查和核对，如有错，应予纠正。

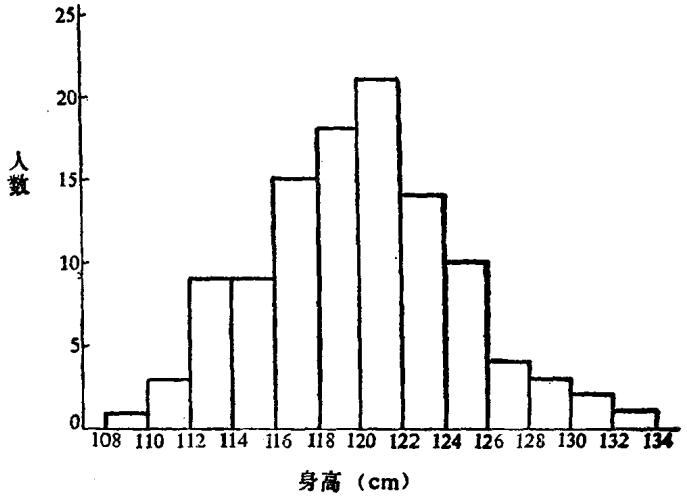

图 2.1　某市 110 名 7 岁男童身高的频数分布

2.2　集中位置的描述

平均数（average）是统计中应用最广泛、最重要的一个指标体系。它常用于描述一组变量值的集中位置，代表其平均水平，或者说它是集中位置的特征值。

平均数的计算和应用必须具备同质基础，必须先合理分组，否则平均数是没有意义的。例如男、女儿童的生长发育规律是不同的，如不分性别地求取某一年龄组儿童的身高或体重平均数，既不能说明男孩，也不能说明女孩的身高或体重的特征，因而是毫无意义的。

常用平均数有均数、几何平均数和中位数，分述如下：

1. **均数**（mean）均数是算术平均数（arithmetic mean）的简称，总体均数用希腊字母 μ 表示，样本均数用 $\bar{X}$ 表示。均数反映一组观察值在数量上的平均水平。

（1）均数的计算方法

1）直接法。即将所有观察值 X_1，X_2，X_3，…，X_n 直接相加再除以观察值的个数 n，写成公式为

$$\bar{X}=\frac{X_1+X_2+X_3+\cdots+X_n}{n}=\frac{\Sigma X}{n} \tag{2.1}$$

式中 Σ 是希腊字母，读作 sigma，为求和的符号。

例 2.2　10 名 7 岁男童体重(kg)分别为：17.3，18.0，19.4，20.6，21.2，21.8，22.5，23.2，24.0，25.5。求平均体重。

$$\bar{X}=\frac{17.3+18.0+19.4+20.6+21.2+21.8+22.5+23.2+24.0+25.5}{10}$$

$$=21.35(\text{kg})$$

2）加权法（weighting method）。当资料中相同观察值的个数较多时，可将相同观

察值的个数，即频数 f，乘以该观察值 X，以代替相同观察值逐个相加。例如表 2.2 的频数表资料，可用各组段的频数作 f，以相应的组中值 (class mid—value) 作 X，按第 (4) 栏求出 fX 及 ΣfX，最后再除以总频数 Σf (即 n)。写成公式为

$$\bar{X}=\frac{f_1X_1+f_2X_2+f_3X_3+\cdots+f_kX_k}{f_1+f_2+f_3+\cdots+f_k}=\frac{\Sigma fX}{\Sigma f} \qquad (2.2)$$

式中 X_1，X_2，X_3，…，X_k 分别为各组段的组中值，即本组段的下限与相邻较大组段的下限相加除以 2，如"108～"组段的组中值 $X_1=(108+110)/2=109$，余仿此。f_1，f_2，…，f_k 分别为各组段的频数。这里的 f 起了"权数"的作用，它权衡了各组中值由于频数不同对均数的影响。即频数多，权数大，作用也大；频数少，权数小，作用也小。故本法称为加权法。

例 2.3　对表 2.1 资料用加权法求平均身高。

$$\bar{X}=\frac{1\times109+3\times111+\cdots+2\times131+1\times133}{1+3+9+\cdots+2+1}=\frac{13194}{110}=119.95$$

7 岁男童的平均身高为 119.95cm。

(2) 均数的两个重要特性

1) 各离均差 (即各观察值 X 与均数 $\bar{X}$ 之差) 的总和等于零。

$$\Sigma(X-\bar{X})=0$$

表 2.2　110名7岁男童身高均数的计算(加权法)

身高组段 (1)	频数, f (2)	组中值, X (3)	fX (4)=(2)(3)
108～	1	109	109
110～	3	111	333
112～	9	113	1017
114～	9	115	1035
116～	15	117	1755
118～	18	119	2142
120～	21	121	2541
122～	14	123	1722
124～	10	125	1250
126～	4	127	508
128～	3	129	387
130～	2	131	262
132～134	1	133	133
合计	110(Σf)		13194(ΣfX)

2) 离均差的平方和小于各观察值 X 与任何数 a (而 $a\neq\bar{X}$) 之差的平方和。

$$\Sigma(X-\bar{X})^2<\Sigma(X-a)^2$$

这两点可用于说明数学上的最小二乘法原理，从上述意义来说，均数是一组观察值最理

想的代表值。

(3) 均数的应用　均数能反映全部观察值的平均数量水平，因而应用甚广。但它最适用于对称分布资料，因为这时均数最能反映分布的集中趋势，位于分布的中心。特别是正态分布资料，均数更有其重要作用，详下章。对于偏态分布资料，均数则不能较好地反映分布的集中趋势，这时需求助于几何均数或中位数。

2. **几何均数** (geometric mean，简记为G)　有些医学资料，如抗体的滴度、细菌计数等，其频数分布明显偏态，各观察值之间常呈倍数变化（等比关系）。这时，宜用几何均数反映其平均增（减）倍数。

(1) 几何均数的计算方法　类似均数的计算，亦可用直接法或加权法，分述如下：

1) 直接法。即直接将n个观察值（X_1，X_2，X_3，…，X_n）的乘积开n次方，写成公式为

$$G=\sqrt[n]{X_1\cdot X_2\cdot X_3\cdots,\ X_n} \tag{2.3}$$

写成对数形式为

$$G=\lg^{-1}\left(\frac{\lg X_1+\lg X_2+\cdots\lg X_n}{n}\right)=\lg^{-1}\left(\frac{\Sigma\lg X}{n}\right) \tag{2.4}$$

例 2.4　5人的血清滴度为：1:2，1:4，1:8，1:16，1:32。求平均滴度。

本例先求平均滴度的倒数，以用几何均数为宜。

$$G=\sqrt[5]{2\times4\times8\times16\times32}=8$$

或
$$\lg G=\frac{\lg2+\lg4+\lg8+\lg16+\lg32}{5}=0.903$$

$$G=\lg^{-1}0.903=8$$

故平均滴度为1:8。

2) 加权法。当资料中相同观察值的个数f（即频数）较多时，比如频数表资料，可用下式计算。

$$G=\lg^{-1}\left(\frac{\Sigma f\lg X}{\Sigma f}\right) \tag{2.5}$$

例2.5　40名麻疹易感儿接种麻疹疫苗后一个月，血凝抑制抗体滴度见表2.3第(1)、(2)栏，求平均滴度。

按式(2.5)用加权法求平均滴度，见表2.3第(3)～(5)栏。

$$\lg G=\frac{\Sigma f\lg X}{\Sigma f}=\frac{72.2471}{40}=1.8062$$

$$G=\lg^{-1}1.8062=64$$

40名麻疹疫苗接种儿童一个月后血凝抑制抗体滴度的平均滴度为1:64。

(2) 几何均数的应用

1) 几何均数常用于等比资料。如抗体的平均滴度和平均效价、卫生事业平均发展速度、人口的几何增长等；或用于对数正态分布资料（详下章）。

2) 观察值不能有0。因为0不能取对数，不能与任何其他数呈倍数关系。

3) 观察值不能同时有正值和负值。若全是负值，计算时可把负号去掉，得出结果后再加上负号。

此为试读，需要完整PDF请访问：www.ertongbook.com