医学统计方法

夏元瑞　主编

人民卫生出版社

医学统计方法

主编　夏元瑞

编者　夏元瑞　杨占清　刘运喜

人民卫生出版社

(京)新登字081号

图书在版编目（CJP）数据

医学统计方法/夏元瑞主编；杨占清等编．-北京：人民卫生出版社，1994

ISBN 7-117-02097-0

Ⅰ．医… Ⅱ．①夏… ②杨… Ⅲ．医学-统计 Ⅳ．R311

中国版本图书馆CIP数据核字（94）第03182号

医学统计方法

夏元瑞 主编

人民卫生出版社出版
（北京市崇文区天坛西里10号）
北京市卫顺印刷厂印刷
新华书店北京发行所发行

787×1092毫米16开本 12印张 286千字
1994年8月第1版 1994年8月第1版第1次印刷
印数：00 001—7 000
ISBN 7-117-02097-0/R·2098 定价：7.50元

编审委员会

序

中国人民解放军济南军区政治部干部部、后勤部卫生部决定在军区开展继续医学教育，组织编写了继续医学教育公共必修课系列教材，《医学统计方法》是其中之一。

医学统计方法是医学科技工作者，在进行科技文献阅读、医学科学技术研究、总结医学经验和建立理论假设等，不可缺少的科学技术手段。因此，要求医学科技工作者，除应该谙熟专业理论知识和技能外，还必须掌握医学统计方法这一科学工具。

《医学统计方法》共十一章，系统地介绍了医学统计方法的基本理论和基本知识。该书对医学统计方法的科学概念有正确而简明的阐述，对统计方法的使用和计算有较详尽的说明，并对统计分析结论的注意事项和使用条件均有较通俗易懂的介绍。通过例题的演算给读者一个清晰的印象和明确的操作程序。

本书对医学统计的常用概念、医学资料的种类、常用医学统计指标、常用医学统计处理方法、抽样研究资料相互比较的显著性检验以及资料的相关分析与回归分析、常用的非参数统计等，均为本书的必读内容，也是正确运用医学统计方法的理论基础。其他如多元回归分析、圆形分布、半数效量等可做为进一步学习提高的部分，在今后专业研究工作中加以选择采用。

该书的编写和出版，给军区医学科技工作者提供了继续教育的必须教材，同时对医学科学研究和经验总结提供了科学的方法。

在此，趁《医学统计方法》出版之际，向编写者表示衷心的祝贺。我非常高兴地将此书推荐给部队医学科技的同仁，并且希望能得到读者的反馈，提出宝贵的建议。

山东医科大学教授　**王均乐**

编写说明

为了促进继续医学教育工作顺利开展，根据济南军区政治部干部部和后勤部卫生部的要求，我们编写了《医学统计方法》一书。

自1978年全国科技大会之后，我国科技人员向科学技术进军已蔚然成风。我军区广大医务工作者也积极地参与科研活动，以提高技术水平和服务质量。为了能透过科研获得的各种数据认识事物的本质，掌握医学统计方法已成为我军区广大医务工作者的急需。本教材较全面地介绍了医学科学研究中常用的一些基本统计方法。全书共十一章，包括常用的统计学指标、显著性检验、相关分析和圆形分布等，基本可以满足各医学专业科研的需要。

近几年来，我军区医学科学研究论文在统计方法应用上常见的问题，是对某些统计指标的概念模糊，致使统计指标选用不当或数据处理公式选用不正确。因而在本教材编写过程中，特别注意了概念的叙述和各种公式的应用条件，力求通俗易懂，使读者容易理解其含义。我们相信本教材对大家会有较大帮助。

在本教材编写过程中，由于水平所限，难免有疏漏或错误之处。诚恳期望广大读者批评指正。

编　者

目　录

第一章 绪　言

第一节　统计学概念和意义

自然界的一切事物都有其内在的规律性，而这种规律性往往被许多表露的偶然现象所掩盖，使我们在观察自然界的事物时，即使在同样条件下，也会出现互相不一致的情况。这种情况就是我们所说的变异性。引起事物变异性的原因不外乎是由于事物自身的变化和观察过程中某些环境因素引起的。不管这种变异的性质如何，它所具有的偶然性都与事物内在的必然性不可分割地相互联系着。科学研究的任务就是从这些表现为偶然性的事物中找出必然性的规律。而科研统计方法则是解决这个问题的一个重要工具。它是以数学中的概率论为根据，对科研实践(动物实验、临床试验和各种调查、观察研究)收集的大量数据进行分析，透过带有偶然性的数据来认识事物的本质，说明各事物间的相互关系，论证事物的发展规律。

医学研究主要以生物为研究对象，其变异因素很多，例如年龄、性别、体重、反应性、环境等。因此医学科研所获得的原始数据都被这些变异因素所造成的偶然性掩盖着。例如某人进行某种药物疗效观察，共观察 40 例，治愈率为 80%，但在同样条件下再治疗 40 例，其治愈率就不一定为 80%，可能大于 80%，也可能小于 80%，这就是个体变异因素的影响。必须用统计学处理才能把握事物的规律性，才能解释两次试验结果差异的内在联系，把自己的科研结论建立在更科学的基础上。另外，目前科学技术飞速发展，知识更新迅速，卫生科技人员必须不断地吸收新的科技信息，需要阅读大量的科研文献，科研统计方法可以帮助读者正确理解文献中某些统计处理的意义，从而推断文献的科学性、可信性和科学价值。所以科研统计方法是广大医学科技人员必须掌握的重要知识和工具。

第二节　统计资料的类型及统计处理方法

一、计 量 资 料

计量资料就是对每个观察单位用定量的方法测定某项指标所得的资料。这类资料大都是用测量工具或仪器测得的。带有度、量或其他单位，所以又叫测量资料。例如身高、体重、血压、脉搏、血色素、白细胞等。对来自这类资料的原始数据的加工是求平均数、标准差和标准误。两均数相比用 t 检验；多均数相比较用 F 检验；两事物关联性分析用直线相关和直线回归分析；总体不明确或开口资料用非参数（秩和检验、中位数检验、等级相关分析等）统计。

二、计 数 资 料

计数资料就是按观察单位的某种特征（或属性）分类，再清点各型中观察单位的个数所得到的资料。例如某些检验结果按阴阳性分开，某些实验动物按生存或死亡分开，

某人群按发病或不发病分开等进行清点。这类资料可计算相对数（比和率）。检验各率间的显著性或相关性可用 χ^2 检验。

三、等级资料

有些资料既具有计数资料特性，又兼有半定量性质，称为半定量资料或等级分组资料。它是将观察单位按某种指标的等级顺序分组，再清点各组观察单位的个数所得的资料。例如某些疾病的一些生化检验，将检查结果按"－"、"＋"、"＋＋"、"＋＋＋"分组，然后清点各组病例数。这种资料具有计数资料的特点，但各组间又有量的差别，它的显著性检验可用非参数统计方法。

第三节　统计中的几个基本概念

一、变异与变量

在医学研究中，存在着两种变异，一种称为个体变异，即观察单位本身的变异，另一种称为随机测量变异。

个体变异：表现为即使条件相同的个体（观察单位）间，各项特征在个体间仍存在着差异。例如，同一年龄的男孩，身高、体重各不相同；同一年龄的成年男子，血清中胆固醇含量各不相同；同一种病的患者，即使病情一致，治疗方法和条件相同，疗效也未必相同等。个体变异的原因在于个体与外界环境关系极为复杂，影响个体特征的内外因素很多，不同因素对于个体的影响偶然地结合在一起，使个体间表现出差异。

随机测量变异：表现为同一个体（观察单位）多次观测结果不完全相同。例如，同一样品，用分析天平多次称量，各次所得测量值不完全相同，我们把这种由于观测手段或条件在种种偶然因素影响下，引起的轻微波动叫做随机测量变异。

由于存在着上述两种变异，所以对某一事物进行观察或研究时，从每个观察单位所取得的某项指标的数值各不相同，我们把这种各不相同的数值称为变量值。

二、总体与样本

总体：是根据研究目的而确定的相同性质、相同类型（同质）事物和现象的某种变量值的全部。例如，欲研究某地区40岁以上正常男子血清胆固醇含量，则该地区所有40岁以上的正常男子的血清胆固醇含量的全部就是总体；欲研究某湖某时期卫生学指标菌的含量，则该湖某时期卫生学指标菌的全部数量就是总体；欲研究某药物疗效，则该药治疗某种疾病总的治疗效果就是总体。

科研的目的就是通过对每个观察单位的调查来了解总体的规律，所以在科研中必须注意在同质的基础上对每个个体的某些特征进行分析和研究，例如要有同一的人口组成；同一的职业年龄分布；同一的诊断标准等。这就是科研统计的总体概念。如果把不同质的资料统计在内，所得出的结论则不科学，不能说明总体的规律。例如"某药治疗Ⅱ期高血压的疗效观察"，把Ⅰ期高血压混在内就可能夸大该药的治疗效果，如果把Ⅲ期高血压混入则有可能缩小治疗效果，均不能正确地表露该药对Ⅱ期高血压的总体治疗情况。所以总体概念是科研的基本概念之一。确定一个科学的总体（也就是科研对象是否同质）

并不是一件容易的事情，要具有较深的专业知识和严格的科学态度才能正确地确定被研究总体的含义。

样本：了解总体情况是统计工作的目的之一。但是多数情况下，总体很大，我们不可能对所有观察单位逐个观察。有时由于观测带有破坏性，如检查一批鸡蛋中沙门氏菌污染情况，即使有可能逐一检查，也不允许这样做。因此我们总是通过对样本的观察去推断总体情况的。

样本就是从总体内随机抽取的一部分。例如，我们从某地抽取100名40岁以上的正常男子，测定其血清中胆固醇含量，这100个胆固醇值就组成了一个样本。为了保证样本能充分代表总体，除样本数量应该满足统计的要求外，样本的抽取必须遵守随机的原则，即要使总体内每一个观察单位都有同等机会落入样本中。随机不是"随意"，它是用某种抽样方法来实现的。例如摸球、抽签、抛硬币等机械随机抽样法或随机表抽样法等（参考《医学科学研究基本方法》继续医学教育公共必修课系列教材）。

三、概率与频率

概率（机率）是事件发生的可能性大小的度量，以符号P表示。可以把我们经常遇到的事件分为三种类型：① 必然事件：指的是必然会发生的事件，如在1大气压下，纯水加热至100℃，必定会发生沸腾现象。必然事件的概率P=1。② 不可能事件：如人可以长生不老事件，必定不会发生，不可能事件的概率P=0。③ 随机事件：指的是在一定条件下可能发生，也可能不发生的事件。如某人在当地流行性出血热流行时，是否会患出血热，回答是不能肯定的，可能患，也可能不患。随机事件的概率P在0与1之间。某事件发生的可能性愈大，则其概率P值愈接近1；某事件发生的可能性愈小，则其概率P值愈接近于0。

频率也是某事件出现的可能性大小的度量。只不过概率是对总体而言，频率是对样本而言。在相同的条件下进行n次重复试验，事件A发生数a(a≤n)。则a与n的比(a/n)为事件A的频率。如n逐渐增大，则事件A的频率就越来越接近概率P。统计学上，常以n充分大时事件A的频率作为该事件概率的近似值。

四、误　差

统计上所说的误差，包括测得值（观察值）与真实值之差和样本指标与总体指标之差。从误差的性质来看，可以把误差分为两大类，即偶然误差和系统误差。

（一）偶然误差　偶然误差包括抽样误差和随机测量误差。

1．抽样误差　科学研究总是从总体中抽出部分（样本）进行实验。用部分的研究结果来说明总体的情况，这种方法叫做抽样调查。由于偶然性的影响，严格按随机抽样调查的结果虽可以代表总体，但不等于总体，两者间总是存在着一定的差异。这种由于抽样所造成的样本与总体或样本与样本之间的差异叫做抽样误差。如某地随机抽取100名40岁以上正常男子血清胆固醇量的均数能代表总体，但不恰好等于某地全部40岁以上正常男子血清胆固醇量的均数，两者存在着差异，这就是抽样误差。由于总体内个体变异必然存在，在抽样调查（或抽样研究）中，抽样误差是不可避免的。

2．随机测量误差　指的是同一个体（观察单位）多次观测结果之差。产生随机测量

误差的原因是观测中存在着的随机测量变异。由于观测中随机测量变异必定存在，因此，随机测量误差也是不可避免的。但是，改善测量手段和测量条件，可以将随机测量误差控制在很小的范围内。

偶然误差的性质：抽样误差与随机测量误差具有共同的性质，即其误差值一般较小；方向是双向的，可正可负；重复观测时其误差的大小和方向不一定重现。偶然误差是由很多影响较小而又难以完全消除的因素综合影响的结果。它的出现是不可避免的，它表面上捉摸不定，实际上却有严格的规律性（如服从正态分布），是可以认识的。科研统计的许多内容就是用概率论的理论透过偶然误差，推论抽样调查结果的可靠性和科学价值。偶然误差（主要指抽样误差）的思想贯穿科研统计的始终。

应该指出，偶然误差是有上述确定含义的统计学概念，不要把它与偶然的误差（偶然发生的误差）相混淆。后者仅仅是指误差发生的方式是偶然的而已。实际上，有些偶然发生的误差却具有系统误差的性质。例如，由于操作的偶然失误，电压的突然变化，仪器的突然故障等所造成的误差，虽然是偶然发生的，却具有系统误差的性质。

（二）系统误差 系统误差是由确定的原因引起的测得值与真实值的较大偏差。产生系统误差的原因很多，其中最常见的原因是试验条件不同。由于试验条件不同而引起的系统误差，又称为条件误差，指的是因试验仪器、试剂、操作方法、诊断方法、治疗方法等条件不同，而造成的结果偏大或偏小。例如，天平砝码未校正，其真实重量（严格地说应为质量）比其所示值偏小时，测得的物质重量将比物质的真实重量为大。在抽样调查中，未遵守同质和随机的原则夸大符合主观愿望的数据，造成的偏差也是科研常见的系统误差。

系统误差的性质：虽然具体的系统误差各种各样，但它们都具有共同的性质。即其误差值与偶然误差相比，一般较大；方向是单向的，或者偏大，或者偏小；在条件不变的情况下观测，同样大小和方向的误差将一再重复出现；具有某个（或几个）明确的原因（虽然这样的原因有时可能尚未发现）；当引起该系统误差的原因消除以后，该系统误差就不再出现；在试验中系统误差会不会出现，出现的大小和方向等并无统计规律性，因此不能用统计方法去认识它。例如，从上面说的天平称量的例子中可以看到，因砝码不准造成的误差是较大的，方向是单一的，即夸大了物质的重量。只要砝码未校正，这样的误差会一再出现。砝码一经校正，这样的误差即消失，其原因是明确的，但当不了解砝码真实的情况时，我们将无法预言它是否会出现以及如果出现，它的大小和方向如何。故科学实验应该消除系统影响因素，从而避免系统误差的出现。

偶然误差与系统误差性质的比较见表 1-1 。

表 1-1 系统误差与偶然误差的性质比较

误差类型	大小	方向	大小和方向的重现性	产生的原因	可否消除	统计规律性
偶然误差	一般较小	双向	不一定重现	是各种偶然因素综合影响的结果	不可避免但可控制	有
系统误差	一般较大	单向	可重现	有少数确定的原因	消除原因即可避免	无

第四节　统计工作中应注意的几个问题

1．常用医学统计的计算随着电子计算机的普及已不成问题，而常见的错误是统计指标或数理统计公式选用的不正确。甚至把最常用的频率指标和构成指标混淆使用而得出错误的结论。产生上述错误的主要原因是对各统计指标和数据处理公式的基本概念理解不深透。所以在学习本统计学时，应把重点放在各种统计方法的基本概念、使用条件和应用要求的理解上。学习和练习计算方法的主要目的也应该是通过例题来理解它的概念和意义，而不深究其数学原理。

2．医用统计方法是解决工作和科研中某些问题的手段，所以它的使用应服从于要解决的主题，根据主题的要求选用适当的统计学指标和统计处理方法，并与实际工作或科研实际情况联系起来分析，说明某种问题才有意义。无目的的单纯为了统计而统计则毫无价值。在文献中有时可见到计算了某项统计指标，却没有联系实际情况的分析，使读者不能领会其价值。

3．统计学中各种数据处理公式都是建立在科学的原始数据基础上的。科学的原始资料是统计处理的关键。一份不准确、不科学的资料（例如系统误差很大），统计处理不但不能解决问题，反而会给人一种错觉，得出某种错误结论。所以，卫生医疗统计资料收集必须持严肃认真和实事求是的态度。科学研究资料必须严格按科研设计方案去收集，例如随机抽样，设立必要的对照并且要有适当的实验例数（参考《医学科学研究方法》继续医学教育公共必修课系列教材），确保统计处理的科学性和可靠性。

第二章　计量资料的统计指标

第一节　计量资料的频数分布

所谓频数就是指观察值的个数，频数分布就是观察值在其所取值的范围内，于各组段中分布的情况。频数分布情况可用频数分布表、频数分布图来表示。编制频数分布表（简称频数表），绘制频数分布图，是整理资料的基本步骤之一，也是对计量资料频数分布情况研究的第一步。通过它可以揭示频数的分布特征、分散的状况和进行分组资料统计指标的换算。

一、频数表的编制方法

例 2-1　以表 2-1 某市 110 名 7 岁男童身高资料为例说明频数表的编制方法。

1．计算全距　找出观察值中最大值和最小值，求它们的差，即全距。本例最大值为 132.5，最小值为 108.2，全距＝132.5－108.2＝24.3。

2．确定组段数、组距和组段　组段数一般取 8～15 个为宜。为分析用，以能显示分布特征为原则；为计算用，应考虑到组段数过多，则计算较繁，过少则误差较大。相邻两组段最小值称组距，各组距可相等，也可不等，一般用等距，如设以组段数为10，则可取全距1/10的整数作为组距。本例全距的1/10为2.4cm，取整数为 2cm，用等距，则共

表 2-1　某市110名男童身高资料(单位cm)

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

表 2-2　110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～	正丅	9
114～	正丅	9
116～	正正丅	15
118～	正正正下	18
120～	正正正正丅	21
122～	正正丅	14
124～	正正	10
126～	丅	4
128～	下	3
130～	丁	2
132～134	一	1
合计		110

分成13个组段。第一组段要包括最小的观察值，组段的下限应低于或等于最小观察值。最后一个组段（应同时写出下限和上限）要包括最大观察值，其上限应包括最大观察值，一般只列出各组段的下限即可。

3．列表归组　根据组段数、组距列出频数分布表(表2-2)。表中第(1)栏为组段。用划记法或分卡法将各观察值分别归入各组段。例如本例所有身高等于及大于108cm，但小于110cm的人都归入"108～"组段内；所有身高等于及大于110cm，但小于112cm的人都应归入"110～"组段内；余仿此进行，得第(2)栏。最后清点各组段内的观察值个数即得各组段频数，得第(3)栏。

二、频数分布图

为了更加直观地了解频数分布情况，通常在编制频数表的基础上，绘制频数分布图。常见的频数分布图为直方图。它以横轴表示观察值，纵轴表示频数。将横轴等分为若干组距为单位的小区间，注明各组段的下限值，在各小区间上作高度等于频数的长方形，即得直方图（如图 2-1 ）。

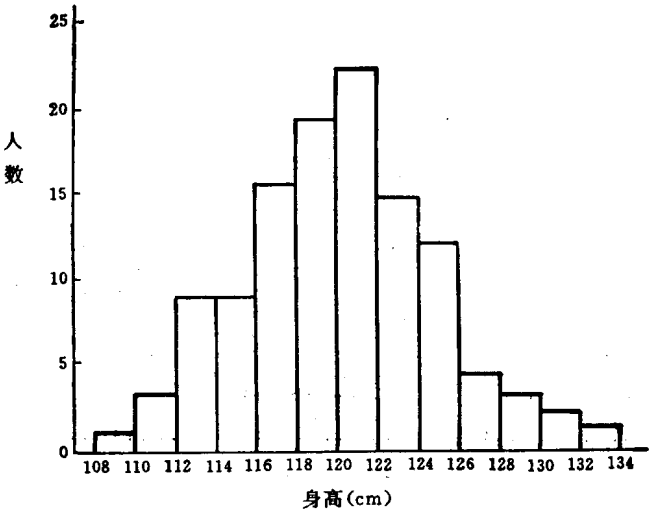

图 2-1　某市110名7岁男童身高的频数分布

第二节　计量资料频数分布的两个基本特征

计量资料频数分布有两个基本特征，即集中趋势和离散趋势。自然界许多事物存在着变异性，但这种变异性是有一定规律的。例如某市 1982 年对 110 名 7 岁男童身高进行了调查，得到 110 个大小不等、杂乱无章的数值，经编制频数表（表 2-2 ）后就可明显看出接近中间数（120～122cm）的频数最多，向两侧逐渐减少。从图 2-1 可清楚地看出频数分布向中部集中形成高峰，即称集中趋势；向两侧逐渐离散形成尾部，即称离散趋势。这种分布形式可分成三种情况：① 高峰在中间，左右两侧大致对称，称为对称分布；② 高峰偏向左侧，右侧尾部延长，称为正偏态分布；③ 高峰偏向右侧，长尾向左侧延伸，称负偏态分布（图 2-2 ）。医学资料多为偏态而且呈正偏态居多。我们研究一群测量资料时，总是要用一定的统计指标来表达这两个特征，便于形成明确的概念。

第三节　集中趋势的指标

平均数是说明同质总体中一群变量值集中趋势的指标。它反映了一群观察值（变量值）的一般水平，并给人们一个简明概括的印象，例如上述 110 名 7 岁男童的身高资料，其值 110 个，数字有大有小，不易形成概念，只有计算出平均数（119.95cm）才能形成明确的印象，即 7 岁男童平均身高在 119.95cm 左右。广义的平均数有 5 种，即算术平均数、中位数、几何均数、众数和调和均数。虽然它们都是表示计量资料集中趋势的指标，

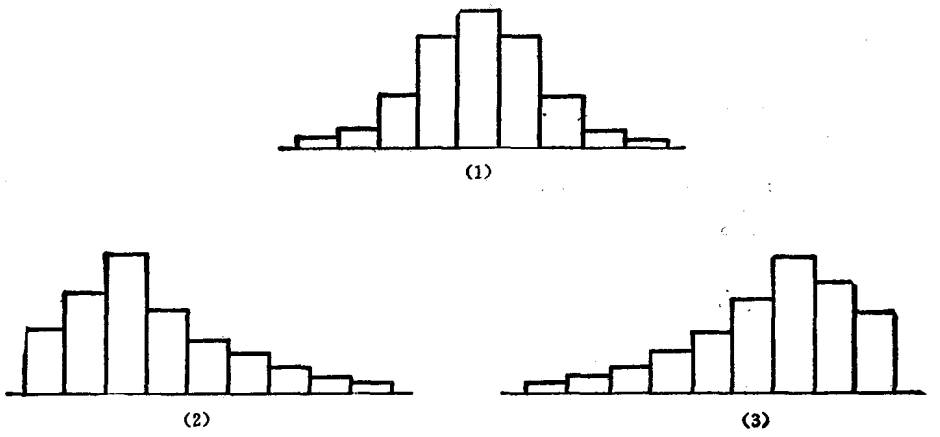

图 2-2　几种不同类型的频数分布示意图

(1) 正态分布型　(2) 正偏态分布型　(3) 负偏态分布型

但其意义不同。为了能够更正确地表达一群变量值的一般平均水平，必须根据资料的分布类型（对称分布或偏态分布）和性质（开口资料、等比数列等）选用。例如统计某种肠道传染病 50 例的住院天数，多数病例住院 8～15 天，平均 10 天。假如其中有少数病例住院超过几十天，那么算术均数必然大于 10 天，因此，不能更好地表达一群变量值的集中趋势，所以在计算均数前，应先对资料进行必要的分析，明确其分布类型，再选定其计算方法。

一、算术均数

(一) 应用条件　算术均数（简称均数）最适用于对称分布，尤其是正态分布资料。因为这时均数位于中央，能反映观察值的集中趋势。当观察值个数较少，而其频数分布基本对称或从专业上可推断总体为正态、近似正态者，也可用均数作为集中趋势指标。

(二) 计算方法

1. 直接法　按式 (2-1) 计算。

$$\bar{X}=\frac{\Sigma X}{n}=\frac{X_1+X_2+X_3+\cdots+X_n}{n} \qquad (2\text{-}1)$$

式中$\bar{X}$为均数；X_1，X_2，X_3，…，X_n 为各观察值；Σ 为求和符号；n 为观察值个数。

例如，有 10 份面粉样品，测得蛋白质含量分别为：10.0，9.5，9.8，8.9，9.2，10.1，9.5，9.7，9.2，9.3g/100 g。求 10 份面粉样品的平均蛋白质含量。

$$\bar{X}=\frac{10.0+9.5+9.8+8.9+9.2+10.1+9.5+9.7+9.2+9.3}{10}$$

$$=9.52(\text{g}/100\text{g})$$

10 份面粉的平均蛋白质含量为 9.52g/100g。

2. 加权法　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，此时观察值应以相应的组中值作 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

$$\bar{X}=\frac{f_1X_1+f_2X_2+f_3X_3+\cdots+f_nX_n}{f_1+f_2+f_3+\cdots+f_n}=\frac{\Sigma fX}{\Sigma f} \qquad (2\text{-}2)$$

式中 X_1，X_2，X_3，…，X_n 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 $X_1=(108+110)/2=109$，其余类推。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2-2 对表 2-2 资料用加权法求平均身高（见表 2-3）。

表 2-3 110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)

按式 (2-2)

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

$$=\frac{13194}{110}$$

$$=119.95$$

7 岁男童的平均身高为 119.95cm。

3. 简捷法　简捷法是将频数表上各组的数值简化成最简单的自然数，再按公式(2-3) 进行运算。此法除计算均数过程比较简单外，对以后计算标准差也很方便。

计算公式

$$\bar{X}=\bar{X}_0+\frac{\Sigma fd}{n}(i) \qquad (2\text{-}3)$$

式中 $\bar{X}$ 为均数，$\bar{X}_0$ 为假设均数，i 为组距，f 为变量值的频数（即个数），d 为差