21st CENTURY 规划教材

21世纪全国成人高等医药院校规划教材

医学统计学

马 彦 主编

中国科学技术出版社

21 世纪全国成人高等医药院校规划教材

医学统计学

主　编　马　彦
编　委　占芬梅　邹红英　赵晓霞　赵凤莲
叶涵泳　吕庆山　王基宁　白翠花
周前明　刘玉波　夏一鑫　韩晓英

中国科学技术出版社
·北京·

21世纪全国成人高等医药院校规划教材
丛书编委会

责任编辑： 周晓慧　高立波
封面设计： 张　磊
责任校对： 刘红岩
责任印制： 王　沛

图书在版编目（CIP）数据

医学统计学/马彦主编．—北京：中国科学技术出版社，2007.7
21世纪全国成人高等医药院校规划教材
ISBN 978-7-5046-4724-5

Ⅰ.医…　Ⅱ.马…　Ⅲ.医学统计　成人教育：高等教育—教材　Ⅳ.R195.1

中国版本图书馆CIP数据核字（2007）第095790号

自2006年4月起本社图书封面均贴有防伪标志，未贴防伪标志的为盗版图书。

出版发行： 中国科学技术出版社
社　　址： 北京市海淀区中关村南大街16号
邮　　编： 100081
电　　话： 010-62103210　**传真：** 010-62183872
印　　刷： 广州市锐先印刷有限公司
开　　本： 787mm×1092mm　1/16
印　　张： 16.75　　**字数：** 240千字
版　　次： 2007年7月第1版
印　　次： 2007年7月第1次印刷
书　　号： ISBN 978-7-5046-4724-5/R·1266
定　　价： 28.00元

前　言

人类与疾病、灾难的斗争历史是永恒的。在与疾病斗争的过程中，预防医学与临床医学各自发挥了不可替代的作用。尤其在突发性公共卫生事件的监测、预警、应急处理中，公共卫生专家和医生们更是控制和消除突发公共卫生事件的危害、保护公众健康和人民生命安全的中坚力量。为了贯彻全国高等医药规划教材研究会、卫生部教材办、全国医科专业第五轮教材会议提出的编写原则与要求，维护和遵从全套教材的整体优化组合，我们组织了一批专家和一些一线教师编写了该教材，以适应21世纪社会与公众日益增长的公共卫生需求。

医学统计学作为整个医学科学的组成部分，是一门与临床医学密切相关的重要课程。本书从卫生统计学的意义、工作步骤及卫生统计学的教与学的方法出发，进行展开讲述。其内容包括数值变量资料的统计描述、正态分布及其应用、总体均数估计和假设检验、方差分析、分类变量资料的统计描述、二项分布及其应用、泊松分布及其应用、秩和检验、直线相关与回归、多元线性回归、Logistic回归分析、常用统计表和统计图、调查研究与设计、实验设计、人口统计、综合评价、医学统计软件应用等内容。

该书强化整体医学思维模式的形成，突出对医学生创新思维的培养，强调创新科学意识与综合素质的教育与训练，给予学生独立思考的空间和机会，促进其辩证医学思维能力的发展。书中涉及内容全面，讲解详略得当，适合作为成人专科学习用书。

在编写本书过程中，由于时间仓促和经验不足，书中难免有一些欠妥和疏漏之处，诚恳希望广大读者批评指正。

编　者

2007年5月

目　录

第一章 绪 论

第一节 学习卫生统计学的意义

统计学(statistics)，是关于数据的收集、整理、分析、解释和表述的科学。统计学分成两个主要领域：数理统计学和应用统计学。数理统计学侧重于建立统计方法和讲述统计方法的原理；应用统计学则是结合特定专业研究特点，使数理统计学原理与方法具体化，从而产生加以特化的统计学，如，社会统计学、心理统计学、生物统计学等。卫生统计学(health statistics)属于应用统计学的范畴，是数理统计学的基本原理和方法在医学，特别是公共卫生学领域的应用，是关于医学、特别是公共卫生研究中资料的收集、整理、分析、解释和表述的一门科学。卫生统计学是进行医学研究中认识事物数量特征与关系的一门方法学，亦是为制定卫生政策提供定量依据的一门方法学。

卫生统计学的主要内容包括：卫生统计学的基本原理和方法、健康统计、卫生服务统计以及调查设计和实验设计等。随着医学科技的进步，学科间的相互渗透以及边缘学科的兴起，数学与电子计算机在医学研究中的应用日益广泛，使数理统计方法成为必不可少的手段。学习卫生统计的目的在于建立统计思维方法，提高科研和工作能力，也是继续教育的需要。

现在，生物医学实验室研究、临床研究、流行病学探索和公共卫生管理都要寻求统计学家的合作。许多医学杂志都邀请统计学家审稿。美国国立卫生研究院(National Institutes of Health，NIH)的基金申请要求合作者有统计学家，并且必须有统计设计与分析的内容。在药物开发中，制药公司要招聘统计学家指导研究设计、分析数据乃至准备报告食品与药物管理局(Food and Drug Administration，FDA)之类机构的文件。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。

第二节 卫生统计工作的步骤

统计是一门科学，它是对观察到的原始数据资料进行加工、解释、作出科学判断的全过程。统计工作的基本步骤如下：

一、实验设计

就是根据研究的目的，制定总的研究方案。包括研究对象的纳入标准和排除标准，样本量和样本获取方法，实验组与对照组的分组原则，确定观察指标及精度，实验过程中的质量控制，拟使用的统计方法，等等。

二、收集资料

收集资料就是根据研究的目的，实验设计的要求，收集准确的、完整的、信息丰富的原始资料。

医学统计资料主要有实验数据和现场调查资料报表、医疗卫生工作记录和报告卡。实验

数据是指在试验过程中获得的数据；现场调查资料主要来源于大规模的流行病学调查所获取的资料；医疗卫生工作记录有门诊病历卡、住院病历卡、化验报告等；报表有卫生工作基本情况年报表、传染病年(月)报表、疫情旬(月、季、年)报表等；报告卡有传染病发病报告卡、出生报告卡、死亡报告卡，等等。

这些资料的收集过程，必须进行质量控制，包括它的统一性、确切性、可重复性。对这些原始数据的精度(precision)和偏性(bias)应有明确的控制范围。

三、整理资料

整理资料(sorting data)　整理资料是把收集到的资料进行适当的分组，把性质相同的资料归纳到一起，用表格或图形的方式展示出来，以反映研究对象的规律性。整理的过程中要核对原始资料的准确性、完整性和可靠性，注意看其是否合乎逻辑，合计是否有误。出现差错要及时纠正，以保证其统计工作的科学性和有效性。

四、分析资料

分析资料就是把经过统计整理的资料，作一系列统计描述和统计推断，阐明事物的规律性。应该注意，不同的资料使用的统计描述和统计推断的方法也会有所不同。

学习统计方法应着重于：

(1)理解医学统计方法的基本原理和基本概念。

(2)掌握收集、整理与分析资料的基本知识与技能。

(3)重视原始资料的完整性、可靠性及处理数据时的实事求是的科学态度。

第三节　卫生统计中的几个基本概念

一、变　量

观察单位是指被观察或测量对象的最基本单位，亦称个体，可以是一个人、一只鼠、一个样品、一个采样点或一个地区等。对每个观察单位的某项特征进行测量或观察，该项特征称为变量，得到的被观察单位的该项特征值称为变量值(value of variable)，亦称作观察值或测量值。例如，某研究其中一项内容是了解某地区2岁以下儿童的卡介苗接种情况，课题组检查了该地区200名2岁以下儿童的卡介苗接种疤痕，这个例子中观察单位为一名2岁以下儿童，变量为卡疤，变量值为“+”或“-”。

定量变量可以分为两种类型：离散型变量(discrete variable)和连续型变量(continuous variable)。离散型变量只能取整数值。例如，一月中的手术病人数，一年里的新生儿数。连续型变量可以取实数轴上的任何数值。有些变量的数值由测量而得到，它们大多属于连续型变量，例如，血压、身高、体重等。“连续”是指该变量可以在实数轴上连续变动。有一些测量值，例如红细胞记数，虽然以“个”为单位时只能取整数值，但其数值很大，当以“千”或“万”为单位时，又可以取小数值，所以通常把这些变量也视为连续型变量。

分类变量(categorical variable)通过确定每个观察单位的某项特征的性质或类别得到的数据，称为分类变量，其变量值是定性的，表现为互不相容的类别或属性，没有度量衡单位。例如，上述介绍的对2岁以下儿童接种卡介苗一例，研究得到的每个儿童卡疤“+”或“-”的数据就是分类变量。通常，作为对分类变量资料进行初步整理，先按类别将观察单位分组，如分为“+”组和“-”组，然后清点每组中的人数，这样得到的数据称为计数资料。

分类变量又可分为几种类型：

1. 无序分类变量 包括：①二项分类变量，特点是其变量值分为两类，如：检查2岁儿童卡疤得到的阳性或阴性；观察某药对某病患者疗效得到的有效或无效。②多项分类变量，特点是其变量值分为两类以上，如职业、血型等变量。

2. 有序分类变量 特点是其变量值是多项分类且各类之间有程度上的差别。如，文化程度可分为：未接受过教育、小学、初中、高中和大专及以上等；疗效可分为治愈、显效、有效和无效。针对这类变量的统计分析方法有秩和检验和等级相关分析等。

变量类型不同与其相适应的统计分析方法也不同。进行处理时要分清是什么类型的资料，选择合适的统计分析方法进行分析。如数值变量资料一般选用，检验或方差分析，分类变量选用 χ^2 检验等。

二、总体与样本

总体是根据研究目的确定的同质观察单位的全体，确切地说，是同质的所有观察单位某种变量值的集合。例如，欲研究某地2005年活产婴儿的出生体重，该地2005年所有活产婴儿的出生体重值就构成一个总体。又如，前文提到的2岁以下儿童接种卡介苗一例的研究，欲了解某地区2005年4月20日2岁以下儿童的卡介苗接种情况，该地该时所有2岁以下儿童的卡疤情况就构成一个总体。这两个例子的总体都明确了一定时间、一定空间，理论上说观察单位的数量是可知的、有限的，称为有限总体。有时总体是抽象的，如欲研究某药治疗胃溃疡的效果，这里总体是指所有胃溃疡病人，但没有时间和地点的限制，观察单位总数量是不可知的，该总体称为无限总体。

对无限总体要观察其中的全部个体是不可能的，即使是有限总体，观察全部个体不仅要花费很大的人力、物力、财力，而且有时也是不必要和不可能的。所以在研究工作中，通常是从总体中随机抽取有代表性的一部分观察单位，我们称其为样本(sample)，用样本信息去推断总体特征。怎样正确地抽取样本，用样本信息去推断总体特征是统计所要解决的问题。

通常，医学研究不可能也没必要对总体中的每个观察单位进行观测或检测，例如，确定某品牌冰棍是否符合卫生标准，只能抽取一定的样本进行检测；再如，欲研究某药治疗胃溃疡的效果，也只能治疗部分病人。通常情况下，医学研究是对样本进行研究，也称为抽样研究(sampling study)，但其目的是通过样本的信息去推论总体的特征。如何用样本信息推论总体，正是卫生统计学的价值之所在。

统计推断的工具是有关概率的理论。如果某事件的结局具有多样性，事先不能肯定，人们就用一系列概率来描述出现各种结局的可能性。既然是推断，既然是由部分推断全体，统计学的结论从来就不是完全肯定或完全否定的。

能不能成功地达到从样本推断总体的目的，关键是抽样的方法、样本的代表性和推断的技术，这些是统计学的核心内容。

三、抽样误差与统计推断

即使在消除了系统误差，并把随机测量误差控制在允许范围内，但是，样本均数(或其他统计量)与总体均数(或其他统计量)之间仍可能有差异，这种误差产生的原因是：①个体之间存在变异；②抽样时只能抽取总体中的一部分作为样本。由此样本的数据构成的统计指标(如均数)就会与总体的该指标有误差，由于这种差异是抽样引起的，故这种误差叫作抽样误差(sampling error)，对它要用统计方法进行正确分析。

一般说来，样本含量越大，则抽样误差越小，样本的观察指标越与总体的该指标接近，即越能说明总体的规律。反之，样本含量越小，则抽样误差相应地越大，因此，我们不能仅

凭观察指标的大小进行简单判断，而应该使用概率与数理统计方法来辨别哪些实验研究的结果是有实际意义的，哪些可能是由抽样误差所造成的，从而得出正确的结论。常用的统计方法是"进行显著性检验"。

进行抽样研究的时候，由于有抽样误差，如果仅仅根据样本指标表面数字的大小就下结论，往往会导致结论错误，只有正确应用统计方法，才能根据样本信息对研究对象的总体规律进行科学的推断，有效地由样本指标估计总体指标；比较可靠地辨别样本指标间的差异仅仅是由抽样误差造成的，还是因为除此之外总体间存在本质差别所致，从而作出较科学的推论。用样本信息推断总体特征叫统计推断(statistical inference)。统计推断包括总体参数估计和假设检验两部分。总体参数估计是按一定的概率估计总体指标在哪个范围，假设检验是首先建立一个关于总体特征的假设，然后根据统计量的抽样分布理论推断检验假设是否成立，并作出统计结论。

四、概　率

概率是描写某一事件发生的可能性大小的一个量度。用 A 表示某一事件，P 表示该事件可能发生的概率，可记为 P(A)。

概率有古典概率与统计概率之分，医学上常用的是统计概率，即对某一随机现象进行大量观察后得到的一个统计百分数。f/N，此处 N 为观察总数，f 为发生数或频数。譬如，某病病死率，乳腺癌术后五年生存率，等等。

在一定条件下，肯定发生的事件称为必然事件，概率为 1；肯定不发生的事件称为不可能事件，概率为 0；可能发生也可能不发生的事件称为随机事件或偶然事件，其概率介于 0 与 1 之间。在统计学上，习惯将 $P\leqslant 0.05$ 或 $P\leqslant 0.01$ 的事件称为小概率事件，表示该事件发生的可能性很小。

如某药治疗 200 个病人，其治愈率为 80%，这 80%是频率。频率是从一次试验或一个样本计算得到的某事件发生率，若经过多次试验或许多人的治疗，其治愈率稳定在 80%，这时可以下结论，某药治愈某病的可能性，即概率为 80%。卫生统计学中的许多结论都是根据概率得到的。一般常将 $P\leqslant 0.05$ 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，是不可能发生的事件，具体的应用在以后的章节中将会介绍。

第四节　卫生统计学的教与学的方法

统计学不是数学，不仅仅如数学那样着重对公理和定理的证明和推导，也不能像学数学那样单纯钻理论、作习题；应用是根本目的，如果不能有效地应用于医药卫生问题，则是最大的失败。统计学不是医学，不能像医学临床课那样要求记忆许多疾病发病原理和细节，也不能像学医学那样事事眼见为实，处处从实际临床病例中总结经验。医学统计学要求在熟悉统计基本原理后，充分理解概念并具体操作实践，对原理和方法的应用才是该学科的根本真谛之所在。

统计学概念与原理并非神秘莫测，它来源于现实生活，要结合生活经验、医学实际安排实施医学统计学的教学内容。例如，摸球模型就是二项分布生动的例子，把黑球和白球看作有病和没病；在同一个口袋里摸球，有人摸到黑球，有人摸到白球，就好像在相同的条件下，有人生病，有人不生病；摸到黑球的总数和生病的总人数是不确定的，但其分布却是有规律的，即二项分布。对于每一个重要的概念和原理，教师和学生都要尽力与一两个实例联

系起来，借助实例来理解一般规律。

正确运用统计方法可以帮助我们科学地认识客观事物，阐明事物固有的规律，但统计决不能创造规律，统计工作最根本的一条是实事求是，必须重视原始资料的完整性、准确性，以严肃认真的科学态度对待数据处理，绝对不允许伪造和篡改统计数字。必须注重应用统计方法解决实际问题，深刻理解各种统计方法的意义及条件，注重结果的解释，对统计公式可不必追究其数学推导。

学习卫生统计学，要掌握群体健康的评价方法，学会运用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，为科学决策服务。

第二章　数值变量资料的统计描述

第一节　数值变量资料的频数表

通过实验或临床观察等各种方式得到的原始资料，如果是计量资料并且观察的例数较多，可以对数据作适当的分组，然后制作频数表或绘制直方图，用以表达数据的分布规律。

一、频数表的编制

所谓频数表(frequency table)是指如下一种格式的统计表：即同时列出观察指标的可能取值区间及其在各区间出现的频数。具体作法：先根据观察个体的数量大小进行分组，然后计算每组中观察值出现的次数。由于这种资料的表达方式较完整地体现了观察值的分布规律，所以也称为频数分布表。

用手工整理资料编制频数表时，通常先编制划记表，即先将选定的组列出，每一组段的起点称下限，终点称上限(上限一般不列出)，然后在原始数据中逐一观察，观察到的数据应当归入哪一组，就在划记表的相应位置上划记，划够五道成一个"正"字。将全部数据划记完毕后计算各组中的笔划数目，即可得到各组的频数。

例 1　抽样调查某地 120 名 18～35 岁健康男性居民血清铁含量(μmol/L)，数据如表 2-1。试编制此血清铁资料的频数分布表。

表 2-1　120 名健康男性居民血清铁含量情况

7.42	8.65	23.02	21.61	21.31	21.46	9.97	22.73	14.94	20.18	21.62
23.07	20.38	8.40	17.32	29.64	19.69	21.69	23.90	17.45	19.08	20.52
24.14	23.77	18.36	23.04	24.22	24.13	21.53	11.09	18.89	18.26	23.29
17.67	15.38	18.61	14.27	17.40	22.55	17.55	16.10	17.98	20.13	21.00
14.56	19.89	19.82	17.48	14.89	18.37	19.50	17.08	18.12	26.02	11.34
13.81	10.25	15.91	15.83	18.54	24.52	19.26	26.13	16.99	18.89	18.46
20.87	17.51	13.12	11.75	17.40	21.36	17.14	13.77	12.50	20.40	20.30
19.38	23.11	12.67	23.02	24.36	25.61	19.53	14.77	14.37	24.75	12.73
17.25	19.09	16.79	17.19	19.32	19.59	19.12	15.31	21.75	19.47	15.51
10.86	27.81	21.65	16.32	20.75	22.11	13.17	17.55	19.26	12.65	18.48
19.83	23.12	19.22	19.22	16.72	27.90	11.74	24.66	14.18	16.52	

直接阅读 120 例血清铁数据，难以产生关于资料分布的明晰印象。将这些数据适当分组，计数每组的频数(例数)。根据这些数据编制成的频数分布表(表 2-2)则能显示出这组数据分布的特点。

表 2-2　120 名健康男性血清铁含量(μmol/L)频数表

组段(1)	频数(2)	频率(3)	累计频数(4)	累计频率(5)
6～	1	0.83	1	0.83
8～	3	2.50	4	3.33

续表

组段(1)	频数(2)	频率(3)	累计频数(4)	累计频率(5)
10～	6	5.00	10	8.33
12～	8	6.67	18	15.00
14～	12	10.00	30	25.00
16～	20	16.67	50	11.67
18～	27	22.50	77	64.17
20～	18	15.00	95	79.17
22～	12	10.00	107	89.17
24～	8	6.67	115	95.83
26～	4	3.33	119	99.17
28～30	1	0.83	120	100.00
合计	120	100		

从表 2-2 可以看出，这组血清铁数据散布在 6～30 μmol/L 之间。频数在各组段之间的分布并不均匀。组段"18～"的频数最多；距离该组段越远，组段的频数越少。

通过如下步骤也可手工编制表 2-2。

(1)找出 120 例血清铁数据的最小值(7.42)与最大值(29.64)。

(2)计算全距(range，R)，也称为极差

$$R=\text{最大值}-\text{最小值}=29.64-7.42=22.22(\mu\text{mol/L})$$

(3)确定组段数与组距：根据数据例数的多少，组段数一般可在 10～15 之间选择。组段的左端点称为下限，右端点称为上限。组距＝上限－下限。实际工作中常按照"组距＝R/(预计的组段数)"的方法估计组距的大致长度。本例如果预计取 12 个组段，则组距长度约为 22.22/12＝1.85。之后，综合考虑以下因素确定本频数表选用的组段数与组距：①两端的组段应分别包含最小值或最大值；②尽量取较整齐的数值(如表 2-2 中的 6、8、10 等)作为组段的端点，这样不仅便于进一步的分析，而且便于对数据进行表述；③组距以相等为宜。

实际上，通过计算机编制频数表时，组距长度的确定也需要研究者干预。

(4)列表：作出如表 2-2 的表格，将选好的组段顺序地列在(1)列。按照"下限≤x<上限"的原则确定每一例数据 x 应归属的组段。依次完成(2)～(5)列的清点频数、计算频率、累计频数与累计频率等步骤，即得到形如表 2-2 的频数表。

在表 2-2 的基础上，可以绘制出图 2-1，称为直方图(频率直方图)。其横轴为血清铁含量，纵轴为频率密度，即频率/组距(直条面积等于相应组段的频率)。在组距相等时，直方图中矩形直条的高度与相应组段的频率成正比。与表 2-2 相

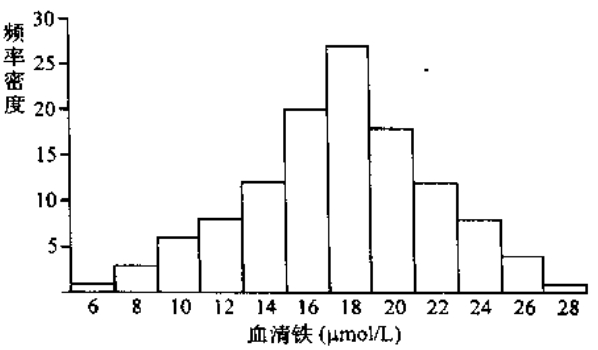

图 2-1　120 例健康成年男子血清铁含量(μmol/L)分布(频率密度＝频率/组距)

比，直方图可以直观地表达出例 1 数据在各组段的分布状况。除去类似表 2-2 提供的信息之外，以组段“18～”为中心，数据分布基本对称的趋势非常明显。许多生物医学变量与例 1 中血清铁类似，分布具有对称的特点。这类分布常被称为对称分布。

例 2　某研究欲了解我国某贫困地区 3 岁以下儿童体格发育现状。2000 年在该地区随机抽取了一定数量的 3 岁以下健康儿童，对其进行了体重、身高、头围等指标的测量。其中 100 名 2 岁组健康男童的身高测量值如下(单位：cm)：

表 2-3　某地区 100 名 2 岁组健康男童身高测量值

75.4	79.5	81.5	83.0	84.0	84.8	85.8	85.2	88.2	89.5
76.3	79.2	81.3	83.0	84.0	84.1	86.0	87.0	88.3	90.1
77.5	79.1	81.2	83.0	84.2	85.0	86.0	87.0	88.3	89.8
78.5	80.2	82.1	83.1	84.2	85.0	86.1	87.1	88.6	90.2
77.6	80.0	82.5	83.1	84.3	85.0	86.3	87.2	88.8	90.5
77.9	80.3	82.6	83.5	84.5	85.2	86.4	87.3	87.6	91.0
79.0	80.8	82.9	83.5	84.7	85.2	86.7	87.5	89.0	91.5
79.6	81.0	82.1	83.5	84.9	85.3	86.7	87.5	89.0	91.2
79.4	81.0	82.4	83.7	84.9	85.3	86.9	88.0	89.3	92.0
79.5	81.3	82.5	83.7	84.8	85.7	86.8	87.9	89.2	93.5

研究者希望通过以上数据对该地区 2 岁组健康男童的身高有一个初步的概括性的了解。

二、用 SPSS 软件绘制频数表

用 SPSS 软件绘制频数表的前期工作，如找最大值、最小值，确定组数，计算组距等，同手工绘制频数表步骤。写组段的工作可以通过 Transform-Recode-Into Different Variables 命令，产生一个新的“组段”变量来实现，然后用 Analyze-Descriptive Statistics Frequencies-Display Frequency Table 命令作出新变量“组段”的频数表。输出结果中 Frequency 为各组频数值，Percent 为各组频数占总数的百分比，Cumulative Percent 为累积百分比。SPSS 软件输出的频数表如下：

表 2-4　组　段

Frequency	Percent	Valid	Percent	Cumulative	Percent
Valid	75～	2	2.0	2.0	2.0
	77～	4	4.0	4.0	6.0
	79～	11	11.0	11.0	17.0
	81～	13	13.0	13.0	30.0
	83～	22	22.0	22.0	52.0
	85～	19	19.0	19.0	71.0
	87～	15	15.0	15.0	86.0
	89～	9	9.0	9.0	95.0
	91～	4	4.0	4.0	99.0
	93～95	1	1.0	1.0	100.0
	Total	100	100.0	100.0	

然后再将上面计算机输出的结果整理成表 2-5 形式的频数表。

表 2-5　2000 年某地区 100 名 2 岁健康男童身高的频数分布

身高组段(cm)	频数	频率(%)	累计频率(%)
(1)	(2)	(3)	(4)
75～	2	2.0	2.0
77～	4	4.0	6.0
79～	11	11.0	17.0
81～	13	13.0	30.0
83～	22	22.0	52.0
85～	19	19.0	71.0
87～	15	15.0	86.0
89～	9	9.0	95.0
91～	4	4.0	99.0
93～95	1	1.0	100.0
合计	100	100.0	

第二节　平均水平的描述

数值变量资料的平均水平，一般反映该组资料的集中位置。其大小用平均数(average)来描述。平均数是一个指标体系。常用的平均数有均数、几何均数和中位数。平均数的计算和应用必须以同质为基础，若把性质不同的资料放在一起计算平均数，则毫无意义。

一、平均数

平均数(average)是描述一组观察值集中位置或平均水平的统计指标，它常作为一组数据的代表值用于分析和进行组间的比较。平均数有多种，常用的有算术均数、几何均数和中位数。

算术均数(mean)简称为均数，用于说明一组观察值的平均水平或集中趋势，是描述计量资料的一种最常用的方法。均数计算有直接法和加权法。

1. 直接法　将所有的观察值 X_1，X_2，…，X_n 直接相加再除以观察例数，写成公式为：

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{\sum X}{n} \tag{2-1}$$

式中，X 表示样本均数；$\sum$是希腊字母(读作 sigma)，为求和的符号；n 为样本观察例数。例如：测量 8 只正常大鼠血清总磷酸性磷酸酶(TACP)含量(μ/L)为 4.20，6.43，2.08，3.45，2.26，4.04，5.42，3.38。试求其算术均数。

按公式(2-1)，算术均数为：

$$\overline{X}=\frac{1}{8}(4.20+6.43+2.08+3.45+2.26+4.04+5.42+3.38)\div 3.9075(\mu/L) \tag{2-2}$$

2. 加权法　当观察值个数较多时，用公式(2-1)计算均数比较麻烦，为计算方便，可先将各观察值分组列成频数表，用加权法求均数。加权法计算公式为：

$$\overline{X}=\frac{f_1x_1+f_2x_2+f_3x_3+\cdots+f_kx_k}{f_1+f_2+f_3+\cdots+f_n}=\frac{\sum fx}{\sum f} \tag{2-3}$$

式中，X_1，X_2，X_3，…，X_n 分别为各组段的组中值，即各组段的下限与相邻较大组段的下限相加除以 2。如“83～”组段的下限为 83，相邻的下一组为“87～”组段，其下限为 85，则