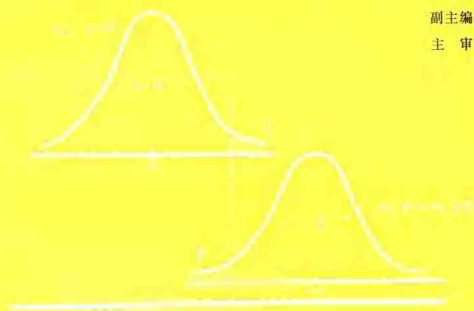


成人高等医学教育用书

HEALTH STATISTICS

卫生统计学

主 编 李明敏
副主编 羊衍惠 许教珊
主 审 杨树勤



5.1

学 术 期 刊 出 版 社

87.9.127

成人高等医学教育用书

卫生统计学

(供卫生、管理、医疗、检验、营养、妇幼专科用)

编写成员按姓氏笔划为序

- 邓名达 (《中国卫生事业管理》杂志社)
王南南 (武汉市职工医学院)
王淑珍 (河北职工医学院)
羊衍惠 (四川省卫生管理干部学院)
许教珊 (山西省职工医学院)
孙维权 (湖北省卫生职工医学院)
刘 骏 (四川省卫生管理干部学院)
李孝儒 (四川省卫生管理干部学院)
李明敏 (四川省卫生管理干部学院)
林中琅 (北京职工医学院)
胡学铭 (武汉市职工医学院)
郭经燕 (河南卫生职工医学院)
彭先导 (湖北省卫生职工医学院)
鞠丽荣 (北京职工医学院)

学术期刊出版社



B 583357

内 容 简 介

本书属成人高等医学教育系列教材之一。全书共十九章，第一章绪论，第二章至第十四章讲述基本统计理论和方法，第十五章至第十九章讲述有关卫生业务统计，即基本统计理论和方法在研究预防医学、临床医学和卫生事业管理等方面的具体应用。

本书主要对象是具有一定实际工作经验和专业知识的在职职工，通过普及卫生统计学的基本原理和方法，以提高其对医药卫生研究资料统计分析的能力。

本书可作为成人高等医学院校的师生及从事卫生、管理、医疗、检验、营养、妇幼保健等专业人员的教学用书和阅读参考书。

成人高等医学教育用书

卫 生 统 计 学

主编 李明敏 副主编 羊衍惠 许敦璐

责任编辑 朱 勤

*

学术期刊出版社出版发行

(北京海淀区学院南路86号)

四川省卫生管理干部学院印刷厂印刷

*

1988年12月第1版 开本 787×1092 1/16

1988年12月第1次印刷 印张: 18.5

印数0001—3000 字数: 528千字

ISBN 7-80045-264-6/R·40

定价: 5.20元



前 言

医学成人教育是我国提高在职医务人员素质和促进我国社会主义卫生事业发展所采取的一项重大措施。各省、市、自治区相继组建了职工医学院校，国家在师资、设备、校舍等方面给予了大力的支持。几年来，通过这些院校培训了大批各种在职的专业人员，为我国的社会主义建设事业作出了应有的贡献。

但是，我国的医学成人教育，在体制、专业设置、教材建设、师资素质等方面还存在不少问题和困难。针对这种情况，一九八六年在太原召开了职工医学教育研讨会，全面讨论了职工医学专业教育的特点、问题和解决办法。与会院校一致倡议编写一套以我国职工医学院教育为基础的，并能供一般医学专科学校和医师进修参考使用的基础教材。这个倡议提出后很快受到卫生部科教司、各省、市、自治区卫生厅(局)及有关医学院校的支持，先后成立了十个教材编写组。经过一年来的努力，这批教材即将相继付印。对于这批教材，总的要求是：密切结合成人教育的特点，确保学员通过学习能够扩大和加深已有的知识基础。为此，教材的编写强调理论结合实际，基础结合临床，在教材内容上要求新颖、精练、实用。但是，限于我们教师队伍的水平，编写时间仓促，客观条件的限制，缺点和错误是难免的，希望同志们批评指正。党的十三大报告中指出：“必须下极大的力量，通过各种途径，加强对劳动者的职业教育和在职继续教育，努力建设起一支素质优良、纪律严明的劳动大军”。这个指示肯定了在职继续教育的作用和任务，也是对我们的鞭策和鼓舞。可以预期，我国的职工医学教育前途光明，任重道远。

职工医学院教材协编委员会

一九八七年十一月

编者的话

医学成人教育是我国整个教育事业的重要组成部分。为促进成人医学教育的发展,我们得到卫生部科教司成人教育处的支持,根据1986年8月在山西太原召开的“职工医学教育研讨会”精神,于1987年4月组成了《卫生统计学》教材编写组。在编写过程中,我们尽量本着使内容在现用高、中等教材的基础上,结合成人教育特点,力求新颖、简明、实用、满足多专业需要的原则,做到内容较为广泛,例题类型较多。但由于时间仓促,书中不足之处在所难免,我们热诚希望使用本教材的广大师生和读者能及时给予指正,以便再版时修订。

卫生部科教司和参加编写的各院校领导对本教材的编写及出版给予了大力支持;特别是华西医科大学杨树勤教授对本教材的审阅提出了宝贵的意见;同时在编写过程中,我们参考和利用了某些作者的资料,在此一并致以衷心的感谢;刘骏老师为本书绘制了大量统计图,谨此致谢。

1988年5月

目 录

第一章 绪论	(1)
第一节 概述	(1)
第二节 统计资料的类型	(2)
第三节 卫生统计工作步骤	(2)
第四节 统计中的几个基本概念	(4)
第二章 集中趋势与离散趋势	(6)
第一节 计量资料的频数表	(6)
第二节 集中趋势	(8)
第三节 离散趋势	(12)
第四节 正态分布及其应用	(16)
第三章 均数的抽样误差及标准误	(19)
第一节 抽样误差的概念	(19)
第二节 标准误的意义及其计算	(20)
第三节 标准误的应用	(21)
第四章 均数的假设检验	(23)
第一节 假设检验的基本概念和一般步骤	(23)
第二节 均数的 t 检验	(24)
第三节 进行假设检验时应注意的问题	(33)
第四节 方差不齐时两样本均数的比较	(34)
第五章 方差分析	(37)
第一节 方差分析的基本思想	(37)
第二节 完全随机设计的多个样本均数比较	(38)
第三节 配伍组设计的多个样本均数比较	(40)
第四节 多个均数间的两两比较	(43)
第五节 多个方差的齐性检验	(45)
第六节 方差分析中的数据变换	(47)
第六章 相对数	(50)
第一节 相对数的概念	(50)
第二节 常用相对数	(50)

第三节 应用相对数的注意事项	(57)
第四节 标准化法	(59)
第七章 二项分布及其应用	(64)
第一节 二项分布的概念及其应用条件	(64)
第二节 二项分布的图形	(66)
第三节 二项分布的均数与标准差	(66)
第四节 二项分布在医学上的应用	(68)
*第八章 Poisson 分布及其应用	(72)
第一节 Poisson 分布的概念及其应用条件	(72)
第二节 属于 Poisson 分布计数资料的分析	(74)
第三节 Poisson 分布曲线的拟合	(77)
第九章 χ^2 检验	(79)
第一节 χ^2 检验的基本思想	(79)
第二节 四格表资料的 χ^2 检验	(80)
第三节 行 $\times$ 列表资料的 χ^2 检验	(83)
第四节 配对计数资料的 χ^2 检验	(86)
*第五节 四格表的确切概率法	(87)
第十章 秩和检验	(91)
第一节 配对比较的符号秩和检验	(91)
第二节 配伍组设计的多个样本比较的秩和检验	(92)
第三节 两样本比较的秩和检验	(94)
第四节 多个样本比较的秩和检验	(97)
第五节 多个样本两两比较的秩和检验	(100)
第六节 秩和检验的优缺点	(101)
第十一章 直线相关与回归	(102)
第一节 直线相关	(102)
第二节 等级相关	(107)
第三节 直线回归	(108)
第四节 直线相关与回归的关系	(114)
第五节 应用相关与回归分析时的注意事项	(115)
第十二章 正常值范围的估计	(117)
第一节 医学正常值范围的意义	(117)
第二节 估计正常值范围的一般原则	(117)

第三节 正常值范围的估计方法	(120)
第十三章 统计表与统计图	(130)
第一节 统计表	(130)
第二节 统计图	(134)
第十四章 调查设计和实验设计	(142)
第一节 调查设计	(142)
第二节 实验设计	(151)
第十五章 卫生事业基本情况统计	(157)
第一节 概述	(157)
第二节 统计报表制度	(157)
第三节 资料的搜集、审核与综合	(160)
第四节 卫生事业基本情况常用统计指标	(162)
第五节 综合分析时应注意的问题	(163)
第十六章 居民健康统计	(165)
第一节 医学人口统计	(165)
第二节 疾病统计	(187)
第十七章 卫生防疫统计	(194)
第一节 疫情统计	(194)
第二节 计划免疫统计	(197)
第三节 卫生监测统计	(201)
第十八章 妇幼保健统计和计划生育统计	(203)
第一节 妇幼保健统计	(203)
第二节 计划生育统计	(207)
第十九章 医院统计	(212)
第一节 医院统计工作的内容	(212)
第二节 医院统计资料的搜集	(212)
第三节 医院统计指标及指标体系	(214)
附 I 统计用表	(239)
附表 1 标准正态分布曲线下的面积	(239)
附表 2 t 界值表	(241)
附表 3 F 界值表(方差齐性检验用)	(242)

附表 4	F 界值表(方差分析用).....	(243)
附表 5	q 界值表 (Newman-Keuls 检验用)	(247)
附表 6	百分率的可信区间.....	(248)
附表 7	Poisson 分布 λ 的可信区间.....	(251)
附表 8	χ^2 界值表.....	(252)
附表 9	T 界值表 (配对比较的符号秩和检验用)	(253)
附表 10	M 界值表 (配伍组比较的秩和检验用)	(254)
附表 11	T 界值表 (两样本比较的秩和检验用)	(255)
附表 12	H 界值表 (三样本比较的秩和检验用)	(257)
附表 13	r 界值表	(258)
附表 14	r_s 界值表.....	(260)
附表 15	百分数与概率单位对照表.....	(261)
附表 16	D 界值表 (正态性检验用)	(263)
附表 17	随机数字表.....	(264)
附表 18	随机排列表 ($n=20$)	(265)
附表 19	两样本率比较时所需样本含量	(266)
附表 20	配对比较 (t 检验) 时所需样本 含量.....	(268)
附表 21	两样本均数比较 (t 检验) 时所需样本 含量.....	(269)
附 II	练习题.....	(270)

第一章 绪 论

第一节 概 述

卫生统计学(health statistics)是把统计的理论和方法应用于居民健康状况研究、医疗卫生实践和医学科学研究的一门应用学科。它是研究医疗卫生各个领域内数据的搜集、整理、分析和推断的科学。医学研究的对象主要是人体及与人体健康有关的因素,由于生物现象的变异较大,各种影响因素错综而复杂,因此仅从所搜集的数据是不能直观地了解事物的本质和规律的。卫生统计学的任务正是从有限的观察中,从表现为偶然的数据中,借助于统计方法,把所研究的事物或现象的本质特征、整体情况和相互关系一一揭示出来,为制定卫生事业计划,评价卫生保健措施的质量和效果;加强卫生事业管理及开展科学研究提供依据。因此卫生统计学不仅是医药卫生科学研究的重要手段,还是实现卫生事业管理现代化不可缺少的工具。

卫生统计学研究的内容主要包括三个方面:

一、卫生统计的基本理论和方法。如医疗和预防科研工作中的调查设计和实验设计;怎样科学、准确而又及时地搜集和整理统计资料,以获得能反映事物实际情况和特征的统计指标;如何运用统计指标,科学地对资料进行分析,通过局部数据来推论全面的统计推断的方法;研究这些方法在医学科学中的应用。

二、居民健康统计。如研究出生、死亡的医学人口统计;研究发病情况的疾病统计,研究人群身体发育水平的发育统计等。

三、卫生事业统计。如反映卫生事业机构、各类卫生人员和床位等,用以研究卫生事业的发展速度、地理分布、比例结构状况、规模及存在问题的卫生事业基本情况统计,反映医院工作质量的医院统计,反映防疫措施和疫情消长的防疫统计等。

根据我国目前成人教育专业设置较复杂的特点,本教材对上述有关内容将分别进行讲解,但重点是卫生统计的基本理论和方法。

电子计算机的发展和普及,对卫生统计学的发展提供了广阔的前景,为资料的搜集、贮存、整理、计算和分析提供了极大的方便,人们正在摆脱繁琐的手工劳动,把卫生统计工作提高到新的水平。

学习卫生统计学的目的是:掌握卫生统计学的基本理论、基本知识和基本技能,为学习各门专业课;阅读专业书刊,从事医疗卫生科研打下必要的统计学基础。在学习时应着重理解统计方法的基本原理和基本概念,掌握搜集、整理与分析资料的基本知识与技能,培养对数据处理严肃认真、实事求是的科学态度;要弄清有关公式的应用条件和用法,不必深究其数学原理;必须联系实际,结合专业知识,进行统计设计和分析,按照资料特点和分析目的,选择适当的统计方法,通过严密的推理,作出合理的判断,培养科学的统计思维方法。

第二节 统计资料的类型

卫生统计资料一般分为计量资料、计数资料和介于二者之间的等级资料，不同类型的资料采用不同的分析方法。

一、计量资料 (measurement data) 是对每个观察单位用定量方法测定某项指标的数值大小所得的资料，一般有度量衡等单位。如身高 (厘米)、体重 (公斤)、浓度 (毫克/升)、脉搏 (次/分)、血压 (千帕斯卡)、白细胞数 (个/立方毫米) 等。这类资料可用第二章、十一、十二章的统计分析方法。

二、计数资料 (enumeration data) 是先将观察单位按某种性质或类别进行分组，然后清点各组观察单位的个数所得的资料。例如对某小学校全体学生进行蛔虫卵粪检，将粪检结果分为阳性与阴性两组，然后清点得两组的人数。又如调查某人群的血型分布，按照 A、B、AB、O 四型分组，清点得该人群各种血型的人数，均属于计数资料。这类资料可用第六章至第九章的统计分析方法。

三、等级资料 (ranked data) 在医学实践中，有些资料具有计数资料的特性，但同时又兼有半定量的性质，我们称它为等级资料。等级资料是将观察单位按某种属性的不同程度分组，再清点各组观察单位的个数所得的资料。例如，检查急性病毒性肝炎病人的血清麝香草酚絮状试验，将检查结果按一、十、++、+++、++++ 等级分组，再清点得到每组人数。又如观察某药治疗某病疗效，将患者按疗效的等级分为痊愈、显效、无效、恶化、死亡五个组，再清点得到各组的例数。这类资料与计数资料不同的是，属性的分组有程度的差别，各组按大小顺序排列，与计量资料不同的是，每个观察单位未确切定量，因而称为半计量资料。等级资料可用第十章的统计分析方法。

资料类型的区分不是绝对的，而是根据研究目的和分析需要来区分的。如白细胞总数原属计量资料，但若按白细胞总数正常 (4000~10000 个/立方毫米) 与异常 (小于 4000 个/立方毫米，大于 10000 个/立方毫米) 分为两组，再清点各组人数，则属计数资料；若按白细胞总数大小分为三个等级，小于 4000 个/立方毫米 (白细胞数减少)；4000~10000 个/立方毫米 (正常)；大于 10000 个/立方毫米 (白细胞数增多)；清点各等级中人数，则属等级资料。

第三节 卫生统计工作的步骤

卫生统计工作包括四个步骤：首先要有一个全过程的设计，然后按照设计的要求去搜集资料、整理资料、分析资料。这四个步骤既有顺序上的先后，又有密切的联系，是前呼后应不能截然分开的整体。即先有科学的、符合实际的统计设计，才能完整、正确、及时地搜集资料；才能整理计算出可靠的统计指标，作出正确的分析结论；而要求分析什么指标，又决定了应搜集哪些资料和怎样搜集整理。故任何一个步骤的缺陷，都会影响统计结果的正确性。

一、统计全过程设计 (design)

统计全过程设计是最关键的一步。主要包括以下内容：

- (一) 明确研究目的。
- (二) 根据研究目的决定搜集哪些资料？但要注意客观条件是否可能办到。
- (三) 用什么方式和方法来取得原始资料？包括拟定适当的调查表格和实验记录。
- (四) 怎样对取得的原始资料作进一步的整理汇总？计算出哪些统计指标？预期会得到什么结果？

为此要求研究者对被研究的事物有一定的了解，结合实际情况，作出科学的、细致的设计，以便用尽可能少的人力、物力和时间，获得准确可靠的结论。

二、搜集资料 (collection of data)

搜集原始资料是根据统计全过程设计的要求进行的。卫生统计资料的来源有两个方面。

(一) 经常性资料搜集。它是结合日常工作来完成的，包括：

1. 医疗卫生工作的原始记录 如门诊记录、住院记录、健康检查记录、卫生检验记录等。它与业务管理和科学研究是紧密结合的，必须认真填写，并注意积累和保存。
2. 报告卡 如出生报告卡、死亡报告卡、急性传染病报告卡、职业病报告卡等。
3. 统计报表 如疫情旬报、月报、季报、医院及防疫工作年报、死因统计报表等。

后两种是国家取得系统的统计资料的一种形式，依据国家规定的报表制度，由医疗卫生防疫机构定期逐级上报。

(二) 一时性资料搜集。它是根据业务或科学研究的需要而临时应用调查或实验方法搜集资料。如为了分析发病规律、探索致病原因、研究生长发育、研究某种药物的疗效、研究地区环境卫生状况而组织的临时性调查等。

搜集资料的要求是完整、准确、及时，这是统计分析准确可靠的基础。完整是指搜集资料项目无遗漏；准确是指观察、测量准确，记录、计算无误；及时主要指经常性资料的搜集，特别是报表应按规定的时间完成。一时性资料的搜集，要求数据的记录应在观察、测量的同时完成，不得以“回忆”方式记录数据。

三、整理资料 (sorting data)

根据统计全过程设计的要求，使原始数据系统化、条理化，便于进一步的计算指标和分析。整理资料的过程包括资料的复核、设计分组、拟整理表和归纳汇总。具体内容和要求见第十四章第一节。

四、分析资料 (analysis of data)

分析资料就是把统计整理的结果，进一步计算相应的统计指标，运用统计方法进行分析对比，对总体作出统计估计，应用假设检验方法进行差异判断，并将统计结果用统计图和统计表加以科学说明，最后，结合专业知识，作出恰如其分的结论。

具体统计分析方法，将在以下各章详述。

第四节 统计中的几个基本概念

一、总体与样本

(一) **总体 (population)** 是根据研究目的确定的同质研究对象的全体。更确切地说,是性质相同的所有观察单位某种变量值的集合。例如研究某地1980年正常人的血压值,则研究对象是该地1980年的正常成人,观察单位是每个人,变量是血压,变量值是测得的血压值,该地1980年全部正常成人的血压值就构成一个总体。它的同质基础是同一地区,同一年份,同为正常成人。

(二) **样本 (sample)** 是总体中随机抽取的一部分。更确切地说,是从总体中随机抽取的部分观察单位的实测值。例如,从某地1980年正常成人中,随机抽取200人,分别测定其血压值,就组成样本。计算样本指标,用样本指标来推论总体指标,这种方法称为抽样研究,这是一种极其重要的科学研究方法。

二、同质与变异

性质相同的事物,就同一观察指标来看,各观察单位(亦称个体)之间,各有差异,称为变异(variation)。例如,研究儿童身体发育,同性别、同年龄儿童(统计上称为“同质”观察单位)的身高,有高有低,各不相同,称为身高的变异。统计研究的是有变异的事物,统计的任务就是在同质分组的基础上,通过对个体变异的研究;透过偶然现象,反映同质事物的本质特征和规律。

三、统计误差

统计上所说的误差,泛指测得值与真值之差,以及样本指标与总体指标之差。误差主要有以下几类:

(一) **系统误差 (systematic error)** 在搜集资料的过程中,由于仪器不准确,标准试剂未经校正、医生掌握疗效的标准偏高或偏低等原因,可使观察结果成倾向性的偏大或偏小,称之为系统误差。这类误差影响原始资料的准确性,应力求避免。如已发生,要尽量查明其原因,予以校正。

(二) **随机测量误差 (random measurement error)** 在搜集资料的过程中,尽管方法已统一,仪器及标准试剂已经校正,但由于各种偶然因素的影响,使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,而是有的稍高,有的稍低,称之为随机测量误差。随机测量误差是不可避免的,但可以努力作到仪器性能及操作方法稳定,例如控制仪器工作条件,熟练测量技巧,使误差降低在允许的范围内。

(三) **抽样误差 (sampling error)** 即使消除了系统误差,并把随机误差控制在允许范围内,但由于总体中各观察单位间存在个体变异,因而样本指标与总体指标之间仍可能有差异,这种样本指标与总体指标的相差称为抽样误差。显然,抽样误差愈小,用样本推断总

体的精确度就愈高，反之亦然。由于生物的个体变异是客观存在的，因而抽样误差是不可避免的，例如居住在同一地区的正常成年男子，他们血中的红细胞数值必定有高有低，因此，如从该地正常成年男子中随机抽取 130 人作样本，则他们的红细胞数（样本均数）和该地全部成年男子红细胞的均数（总体均数）不一定相等，这就出现了抽样误差。但抽样误差是可以控制的，抽样误差有一定的规律性，运用这些规律可以进行总体估计，详见第三章。

四、概率与频率

（一）**概率**（probability）是描述某事件发生的可能性大小的一个度量。以符号 P 表示。在一定条件下，肯定发生的事件称为必然事件，其概率 $P=1$ 。肯定不发生的事件称为不可能事件，其概率 $P=0$ 。在一定条件下，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。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 0。

（二）**频率**（relative frequency）也是指某事件发生的可能性大小的一个度量。不过概率是对总体而言，而频率是对样本而言。即在相同条件下，进行 n 次重复试验，事件 A 发生数为 f ($f \leq n$)，则 f 与 n 的比 f/n 为事件 A 的频率，随着 n 的增大，事件 A 的频率越来越接近概率 P ，故当 n 充分大时，事件 A 的频率可作为事件概率的近似值。

五、参数与统计量

（一）**参数**（parameter）是用来描述总体特征的指标，如总体均数、总体率、总体变异指标等。参数常用希腊字母来代表，如总体均数为 μ ，总体率为 π ，总体标准差为 σ 等。

（二）**统计量**（statistic）是由样本计算的相应的指标，例如样本均数、样本率、样本标准差等，统计量常用拉丁字母来代表，如样本均数为 $\bar{X}$ ，样本率为 p ，样本标准差为 S 等。

复习思考题

1. 卫生统计学的意义是什么？它有哪些任务和内容？
2. 卫生统计工作有哪些步骤？各个步骤基本内容和要求是什么？
3. 统计资料有哪些类型？各举实例说明。
4. 观察值的变异由什么原因引起，试举例说明。
5. 什么是抽样研究方法？什么叫抽样误差？

（计教编）

第二章 集中趋势与离散趋势

第一节 计量资料的频数表

一、频数表的概念

表2-1中第(1)栏的“每家患者数”为各观察单位(每个家庭)的观察值,第(2)栏的“家庭数”为相应的观察单位数,亦称频数。列有类似第(1)、(2)栏数字的统计表称为频数表,亦称频数分布表。

表2-1 每家某病患者数的分布

每家患者数	家庭数
0	20
1	80
2	40
3	50
4	10
5	10

二、频数表的编制方法

例2-1 某地区130名正常成年男子红细胞数(万/立方毫米)资料如下,试编制频数表。

379*	457	519	486	428	467	537	498	445	588**
453	516	484	415	460	531	497	443	477	478
510	483	411	463	528	494	440	474	567	505
481	398	461	523	490	435	470	546	503	449
389	457	521	487	429	467	538	498	446	478
454	516	485	417	466	532	497	443	477	507
513	483	413	464	529	495	442	474	569	453
481	401	462	526	491	436	473	549	504	478
394	457	523	490	431	468	539	499	443	508
454	517	486	427	466	536	498	443	477	453
515	484	413	464	529	496	442	475	569	480
482	410	462	526	493	439	474	561	504	510
398	458	523	490	433	468	540	500	449	480

*最小值 **最大值

(一) 计算全距

找出观察值中的最大值与最小值，求它们的差，即全距 (range)，亦称极差，常用R表示。本例最大值为588，最小值为379，即

$$R = 588 - 379 = 209$$

(二) 确定组段数，组距和组段

组段数一般取8~15个，过多计算较繁，过少误差较大。根据组段数和全距，决定组距，常用*i*表示。

$$\text{组距} = \frac{\text{全距}}{\text{组段数}}$$

本例全距为209，组段数暂定10，组距为 $\frac{209}{10} = 20.9$ ，为计算方便，取整数为20。

划分组段时要注意，第一组段要包括最小观察值，最末组段要包括最大值，故最后定组段数为11。划分组段常用表2-2第(1)栏的表示法，如第一组段370~，第二组段390~，…，第十一组段570~590。

(三) 列表划记

如本例，可列出表2-2的划记表，再将原始资料用划记法计数，得各个组段的频数，表中第(1)、(3)栏即为频数表。

表2-2 某地区130名正常成年男子红细胞数划记表

红细胞数组段(万/立方毫米)	划 记	频 数
(1)	(2)	(3)
370~	丁	2
390~	正	4
410~	正正	9
430~	正正正	16
450~	正正正正丁	22
470~	正正正正正	25
490~	正正正正正	21
510~	正正正正丁	17
530~	正正	9
550~	正	4
570~590	丁	1
合计		130

三、频数表的用途

(一) 揭示频数的分布特征

频数表可以直观看出频数分布有两个重要的特征，即集中趋势与离散趋势。从表2-2中，可以看出，红细胞数过多或过少的人是少数，而较多的人集中在中间组段上(即470~组

段),这种特征称为集中趋势。另外,也可看出红细胞的数值是参差不齐的,这种参差不齐的程度称为离散趋势。这两个重要的特征是从两个不同的角度揭示计量资料的频数分布规律。

(二) 揭示频数的分布类型

将表 2-2 绘成图 2-1, 可看出频数分布的对称性。图 2-1 高峰位于中部, 左右两侧的频数大体对称, 为对称分布。在这类分布中, 最典型的是正态分布。医学领域中常见的频数分布除对称分布外, 还有正偏态分布(高峰偏于左侧, 长尾向右侧观察值较大的一侧伸延)及负偏态分布(高峰偏于右侧, 长尾向左侧观察值较小的一侧伸延)。医学资料多为偏态分布, 而且正偏态居多。了解频数分布类型, 以便选择适当的统计方法。正态分布或近似正态分布的资料可按正态分布原理作统计处理, 偏态分布资料可用非参数法作统计处理。

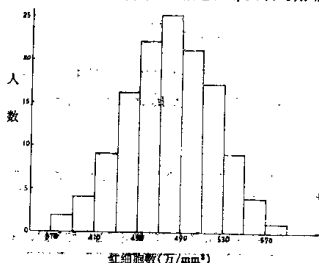


图2-1 某地130名正常成年男子红细胞数频数分布

(三) 便于发现某些特大或特小的可疑值。必要时经统计检验决定取舍。

第二节 集中趋势

定量描述集中趋势的指标是平均数, 它是统计中应用最广泛、最重要的一个指标体系, 它反映一组观察值的平均水平或典型水平。由于原始数据的分布不同, 因而它的平均水平可用各种不同的平均数表示, 常用的平均数有算术均数、几何均数和中位数。

一、算术均数

算术均数 (arithmetic mean) 简称均数, 当所处理资料的观察值大小分布比较对称时, 可用算术均数。总体均数用希腊字母 μ 表示, 样本均数用 $\bar{X}$ 表示。

(一) 均数的计算

1. 直接法 将所有观察值 $X_1, X_2, X_3, \dots, X_n$ 直接相加再除以观察值的个数 n , 写成公式为

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n} = \frac{\sum X}{n} \quad (2.1)$$

式中 Σ 是希腊字母, 为求和的符号。