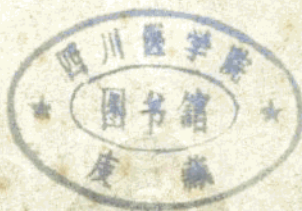


館存

卫生统计学讲义



125054
#

湖南省卫生防疫站编印

一九七六年五月

前 言

在毛主席革命卫生路线的指引下，我省卫生防疫工作正在逐步加强。随着形势的发展，各地卫生防疫人员为了进一步提高卫生、流行病学和科学实验工作质量，迫切要求掌握卫生统计方面的知识。为此，我们参考有关资料，结合我们工作中的体会，编写这本讲义，其中部分章节计算举例录用原参考材料。

这本讲义介绍了卫生统计工作有关资料收集整理、图表制作，指标运算和科学研究中常用的统计方法，希望对各地卫生防疫工作同志在实际工作中有所帮助。

由于我们学习不够，水平有限，在编写上仍存在不少缺点和错误，请读者批评指正。

毛主席语录

把医疗卫生工作的重点放到农村去。

应当积极地预防和医治人民的疾病，推广人民的医药卫生事业。

人的正确思想，只能从社会实践中来，只能从社会的生产斗争、阶级斗争和科学实验这三项实践中来。

胸中有“数”。这是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的

数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无“数”，结果就不能不犯错误。

应当从客观存在着的实际事物出发，从其中引出规律，作为我们行动的向导。为此目的，就要象马克思所说的详细地占有材料，加以科学的分析和综合的研究。

卫生统计学讲义

目 录

第一章 概述

- 第一节 卫生统计学的性质及研究对象..... (1)
- 第二节 卫生统计学的任务和要求..... (2)
- 第三节 统计工作中的几个基本概念..... (2)
- 第四节 统计工作的步骤..... (3)

第二章 资料的收集和整理

- 第一节 数量标识..... (5)
- 第二节 资料的收集..... (5)
- 第三节 资料的整理..... (6)
- 第四节 频数分配..... (6)
- 第五节 统计表..... (9)
- 第六节 图示法..... (12)

第三章 均数及变异指标

- 第一节 均数的意义..... (21)
- 第二节 均数的计算..... (21)
- 第三节 指标的变异性..... (27)
- 第四节 全距及均差..... (27)
- 第五节 标准差..... (29)
- 第六节 变异系数..... (33)

第四章 相对数

- 第一节 绝对数和相对数..... (35)
- 第二节 相对数的计算和运用..... (35)

第三节	动态数列和动态指标	(39)
第四节	相对数的平均数	(42)
第五节	标准化法	(42)

第五章 抽样问题、显著性测验

第一节	常态分配	(44)
第二节	抽样	(46)
第三节	标准误	(48)
第四节	显著性测验的意义	(50)
第五节	t 分配及 t 测验	(50)
第六节	相对数的显著性测验	(54)
附录	t 分配表	(56)

第六章 X^2 测验

第一节	X^2 基本计算方法	(57)
第二节	四格表	(59)
第三节	$2 \times K$ 表	(61)
第四节	$R \times C$ 表	(62)
附录	X^2 值表	(64)

第七章 几何平均数、中位数

第一节	几何平均数的运算	(65)
第二节	几何平均数的标准差	(66)
第三节	几何平均数的显著性测验	(67)
第四节	中位数	(69)

第八章 半数致死量 LD_{50}

第一节	半数致死量的意义	(72)
第二节	累计法	(72)
第三节	累计图解法	(75)
第四节	机率单位法	(76)
附录一	计算 LD_{50} 时距离比查对表	(79)

附录二	死亡率转换为机率单位表	(81)
第九章 相关		
第一节	相关及相关系数	(82)
第二节	小样本相关系数的计算	(83)
第三节	大样本相关系数的计算	(84)
第四节	相关显著性测定、相关系数的合并	(86)
第五节	等级相关 (相关顺序测验)	(90)
附录一	相关系数之 5%界与 1%界	(93)
附录二	r与Z对照表	(94)
附录三	Spearman 顺序相关显著界	(94)
附录四	相关顺序测验显著界	(95)
第十章 回归		
第一节	直线回归	(96)
第二节	回归系数的显著性	(101)
第三节	曲线回归	(103)
第十一章 非参数统计		
第一节	符号检验 (优势检验)	(108)
第二节	两样本的符号顺序检验	(110)
第三节	单因素多组的顺序检验	(113)
第四节	两因素多组顺序检验	(115)
第五节	参照单位分析法	(117)
附录一	Wilcoxon 符号顺序检验显著界	(120)
附录二	顺序和H值显著界, 双侧检验 $n_1 \leq n_2$	(121)
附录三	三组数据顺序检验显著界	(122)
附录四	二因素多组顺序检验的显著界	(122)
第十二章 实验设计		
第一节	实验设计的意义和步骤	(123)
第二节	样本	(124)

第三节 处理及处理分配	(127)
第四节 人群防病效果观察设计应注意问题	(131)
第五节 分析	(131)
附录一 两组百分比相差显著时所需例数	(132)
附录二 两组百分比相差非常显著时所需例数	(133)
附录三 随机数目表	(134)
常用对数表	(136)
反对数表	(139)
平方表	(142)
平方根表	(145)
常用数学符号表	(150)
常用计量单位换算表	(150)

卫生统计学讲义

第一章 概 述

伟大领袖毛主席教导我们：“胸中有‘数’。这是说，对情况和问题一定要注意到它们的数量方面，要有基本的数量的分析。任何质量都表现为一定的数量，没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面，不懂得注意基本的统计、主要的百分比，不懂得注意决定事物质量的数量界限，一切都是胸中无‘数’，结果就不能不犯错误。”（党委会的工作方法）指出了事物的发展过程 and 变化，必须用数量表明，才能具体、明确、一目了然。要了解事物的数量变化，就要运用统计学方法。

统计学是社会主义经济建设、文化建设中的必要学科之一，目前卫生部门的各专业在业务开展、调查研究及经验总结工作中都日益重视使用统计学知识。卫生防疫部门在贯彻“预防为主”，调查疾病规律，分析防治效果，研究、落实预防及各项卫生措施方面，都日益广泛地使用基本的统计学方法。

第一节 卫生统计学的性质及研究对象

统计学是研究社会生活的科学，属于社会科学的范畴。它的重要特点是：在研究社会生活的具体现象和发展过程时，用数量来说明所出现现象的普遍程度、发展趋势、速度及其相互关系，用数量来表示质的变化的特征及其发展过程的规律性。

统计学是以数学理论为基础的，统计方法是用通过数学运算所得出的计算公式，合理使用于各项业务。统计研究是以大量同类现象为对象，应用特定的方法收集这些现象，然后将所得资料加以整理和分析，找出代表现象总体特征的数量指标，说明规律，以指导实践。

统计学作为一门独立学科，是普通统计学理论，称为数理统计学，它的研究对象是一般统计方法及其理论根据。但统计学的研究更重要的是应用于各种业务部门。由于各业务部门各有其不同的特点，它所表现的现象及发展过程各有其特殊性，因而在研究过程中亦有不同的特点。卫生统计学是根据卫生部门的特点，运用统计方法研究有关现象的发展过程中质和量的变化及其规律性，同时也研究人民健康水平。使医疗卫生工作更好地为提高人民健康水平而发挥积极的作用。

目前卫生统计学已广泛的应用于基础医学、临床医学和预防医学等各方面的自然科学的问题研究。在卫生防疫工作中，卫生学、流行病学及微生物、寄生虫学等方面的研究，卫生统计方法的应用也已成为不可缺少的内容。由于自然科学研究的对象是生物，生物

有变异特性,研究过程中须考虑其变异规律,因而在统计方法的应用时更加深入和细致。

第二节 卫生统计学的任务和要求

卫生统计资料是卫生事业发展的依据,卫生统计方法又是卫生科学研究的必要方法。因此卫生统计学就有以下的任务和要求:

首先,统计必须为社会主义建设服务,要有严格的科学性。它的服务对象是有阶级性的。在资本主义国家里,统计是为资产阶级服务的,他们为了垄断资本的利益,对现代资本主义发展的真面目进行了无耻的粉饰,千方百计的利用不正确的统计方法,不断制造谎言,企图掩盖其腐朽的本质。我们必须揭露和批判资产阶级违反科学原则的统计方法。

我们的统计学是为社会主义建设服务的,是按照严格的科学原则的,具有高度的可靠性和精确性。卫生统计学工作,必须认真执行毛主席的无产阶级革命卫生路线,在**预防为主**的方针指导下,不断总结实践经验,运用于实践。

其次,统计必须为计划服务。为了使卫生业务计划、医学科学研究计划能符合客观实际,在制订计划时,必须有科学依据,因而须掌握正确可靠的统计资料。如制订门诊工作计划,医院病床设备及医务人员配备计划,就须掌握当地人口、生产、生活及一般发病率水平的情况;在制订年度或季度卫生防疫工作计划时,必须掌握当地生产活动、自然条件、城乡人群分布及各种季节性、地方性疾病流行病学资料等情况。

计划的制订仅仅是工作的开始,还要组织人力、物力,采取正确措施,以争取计划的胜利完成。在计划执行过程中,要运用统计指标检查计划执行的质和量的水平。正确的统计指标能反映合理的工作方法和存在的问题,从而可以更好地发挥积极因素,推广先进,改正缺点,以提高工作质量。

第三,统计研究必须正确使用统计方法。卫生服务工作的效果反映在人民的健康水平上。为了正确反映人民健康水平,要求统计工作提供可靠的数字资料,使资料分析能正确地反映卫生工作在保证人民健康、提高健康水平上的成绩及其在经济建设中所起的作用。同时统计资料也应如实指出存在的问题并提出可行的解决办法。为达到这个任务,在资料分析过程中必须正确运用统计方法。

医学科学研究的成果须有统计资料作依据,临床研究,流行病学、卫生学及实验室研究均须有合理的统计设计,并正确运用统计方法进行归纳和分析,才能获得正确的结论。

第三节 统计工作中的几个基本概念

事物的现象或发展过程是遵循一定的规律进行的。而表现事物出现的现象或过程的有关个别事例是众多的。为了正确认识事物发展过程的规律性,必须对所出现的有关的、众多的个别事例进行收集整理和适当的分析处理。在分析研究的各阶段中,应自始至终以获得事物现象的全部及其有关个别事例所集中的特征(绝对值、比值及均数)为目的。因而在统计研究时,必须明确理解以下几个基本概念:

一、总体概念

为了正确了解事物的现象，进行统计研究时对所有有关的应进行观察的事例都必须在考虑之内，而无关的事例，则不应混淆其中。在所有有关事例中，凡同质性的事例结合而成的集团，成为总体。例如在人口统计时，某地某时的居民就是一个总体；而各个居民则是这个总体中的一个单位，他们都具有许多不同的特点或标识，如性别、年龄、职业、生活习惯等。同样，某地某时某病的患者或死亡均可分别视为同质性的总体。

总体是客观存在的，它是以研究的目的予以决定的。在统计研究中必须具有总体概念。

二、变异概念

在同质基础上构成的总体，其中各单位在性质上或数值上是有变化的，即其构成单位的标识是具有变异性的，这种变异标志可以用文字来表示质的变异，也可以用数字表示其量的变异。例如我们常用均数或比例数来代表总体的特征，但这些指标是从各观察单位收集的资料得出的。而各观察单位所表现是不一致的，互有差异的，作为典型特征的代表值就必然具有变异性，如说某病平均病程若干天，但各病例的病程是有长有短的，有一个伸缩的幅度的，这就是变异的幅度或范围。所以，为了细致地深入地了解总体，叙述总体，说明总体，确实有效地发挥统计学指导实践的作用，在统计研究中必须具有变异概念。

三、大数量概念

同类事物的各个单位不是完全相同的，有一定的差异的，其一般规律表现在集体中，表现在大量事例的总体中。我们研究事物现象的发展过程通常是从总体中抽取代表性的事例来分析、说明总体，统计学研究正是从事物总体中及其联系中抽取代表性事例的武器。

我们研究事物发展过程的规律性，常不能由某些少数个别事例中找出来，必须从大量的观察中掌握丰富的资料，从各个现象之间的相互联系、相互依存关系中进行研究，始能找出规律。为了避免对事物认识的片面性，使对现象的了解尽可能地接近真实情况，就须有大量的材料，这就须进行较大数量的观察。因此，如实地说明总体及其构成，应有大数量概念。同时必须注意：大量观察资料之科学加工，是发现规律的手段，是表现规律的形式，而现象本身的发展规律，绝不是大量观察所决定的。大数量观察所获得的指标更接近事物客观存在的真实规律。

总之，为了了解现象，必须确定总体；为了深入细致地了解总体，必须掌握并理解变异概念；为了使统计指标更接近事物的真实现象，必须具有大数量概念，这三个基本概念又是相互联系，都是为正确认识现象所不可缺少的。

第四节 统计工作的步骤

进行统计研究时，首先要明确研究的目的，确定须用的统计方法说明其主要问题的

内容。因此，在实际工作中进行统计研究时必须全面地熟悉所要研究的问题。例如，运用统计方法研究流行病学时，必须了解急性传染病的特点、流行规律及其防治方法，才能提出切合实际的研究目的和要求。

目的明确后，要进行统计研究设计。根据目的和要求，设计具体的统计观察方法，包括如何收集正确完整的资料和对资料进行整理分析的有关问题，如确定计算指标，资料收集方法等。资料收集后再运用统计方法进行分类、分组计算，求出正确的指标。

最后，还要按原设计的目的要求及所求得的各项指标，运用历史唯物主义和辩证唯物主义的观点和方法，对所研究的事物现象，进行科学的阐释，以求得正确的结论。

此外，在统计计算过程中常常应用基本数学，特别是运用代数、对数运算方法，或应用珠算、计算尺、计算机等工具来运算、对这些计算方法和运算工具的使用应掌握纯熟。

第二章 资料的收集和整理

第一节 数量标识

事物的总体是由同性质的各单位构成的，总体中各单位标识在性质上、数值上是有差别的，这些有差别的各个标识，称为变异标识。变异标识有两类：事物现象在发展过程中所出现的量的变化，称为数量标识；质的变化，称为品质标识，数量标识是可以用数字表明的，而品质标识则不能单纯用数字表明，需加以文字说明，这两类变异标识也不是绝对划分的，它们是相互联系，可以互相转化的。

数量标识亦称变量，表示某一具体事物数量标识的具体数值称为标识值，或称变量值，亦即读数。医学上各种数字资料中的数量标识根据其来源可分为两大类：一类是测量所得的记录，如身高、体重、脉搏、血压、体温、红、白血球数，以及时间、距离等等，叫做测量资料；另一类是清点数目所得记录，如疫苗接种反应人数、寄生虫病检查虫卵携带者阳性或阴性人数，带菌者调查人数，各种传染病发病及死亡人数等等，叫做计数资料。

观察事物现象的数字资料所得的读数，常常是很多，而且常无一定次序。由于这些记录的变异性，我们在了解它们所包含的意义，以及使和同类的另一批记录进行比较时，就会感到困难。为了解决这个困难，并对所研究的现象获得正确的认识，我们就须将收集的资料，加工整理，并用统计学方法，进行运算。

第二节 资料的收集

统计资料的收集，应依据统计研究设计的要求，有关项目力求完整，各观察单位均须在所要研究的总体范围内，无关的资料应予删除，避免混淆不清。

在医学研究中资料来源及收集的方法有以下各方面：

1. 从有关部门取得基本资料：如流行病学调查分析中所需的人口资料、气象资料、野生动物资料等可向当地有关部门联系取用。为了适应某种特殊分析的要求，也须通过调查收集，如基本人口资料中的年龄、性别分布等。

2. 流行病学调查方法：根据研究目的，确定调查对象、范围和內容，编制调查纲目及调查表，组织一定的力量进行直接调查，有些也可采用通讯调查方式。调查表设计时要注意每一项目或指征要有一致的标准。记录必须清楚，不能含混。

3. 实验室记录，使用临床检验或实验室资料时，设计时应要求实验室记录清晰，将结果抄录，再累计于整理表上。

4. 从临床病案中收集：对某些疾病的发病率、病死率的研究，药物治疗效果的研究，均可从临床病案中收集资料，应先设计登记表，将病案中有关项目记录在登记表中，然

后进行统计分析。

此外，还可应用传染病登记、报告资料及其他有关资料进行相应的分析研究。

第三节 资料的整理

统计研究工作应将收集的资料，进行加工整理，以同类现象为基础，使之系统化，并找出其典型的特征，便于科学的分析。

一、资料整理的步骤

1. 审查原始资料：对所收集的各项资料作初步审查核对，尽可能使符合设计要求，注意资料登记的完整性和正确性。不在研究范围内的资料及重复的资料予以删除；缺项的或记录模糊的可进行补充调查补正；无法补正的或缺项过多而对总体研究影响不大的资料可以剔除。

2. 分类、分组累计成整理表：为将有关资料中的同类现象进行累计，应设计整理表，予以分类或分组累计。如按行政区划分，可按公社、大队或生产队等队别；按地理划分有山区、平原、丘陵等；按属性划分有性别、职业、病种等；按数量划分有身长、体重、年龄等；流行病学中还有按病原型的划分，实验室中还有按血清滴度的划分等等。将各类、组出现的次（例）数予以累计，制成整理表。

3. 使用统计方法运算，求得各项指标（方法见以后各章）

4. 制成统计表及必要时使用图示法。

5. 运用计算所得的指标来说明结论。

二、分组归纳

将原始资料中各数值进行分类、分组整理时，可采用以下方法：

（1）划记法：将应归纳于各组的个数划记标明，通常应用以下符号，每笔代表一个标识值：

一 丁 下 正 正 | || ||| |||| 册

这种方法简便、易用。但如疏忽，易于发生错误，错了又须从头记起。适用于分组简单、资料较少的情况。

（2）分卡法：即将记录标识值的卡片，按不同的组分开来，然后清点每组的卡片数，记入整理表内。

分组划记或分卡后，将同类、同组各数值的个数分别累计，累计时可直接计数，或用算盘、计数器等辅助计算，分别将结果记入整理表的相应各栏内。

（3）边缘打洞卡。（略）

第四节 频数分配

在统计工作中，对连续性的资料如身长、体重、年龄、时间的分组统计时，常用频

数分配方法。

频数分配是一种特殊的统计表，它表明一套标识值的分组及归纳于各组中的标识值的个数。从统计研究的数学分析的观点来说，它是一种最重要的表式。我们把原始的标识值加以有系统的分组，使之成为紧凑而简略的结构，易于获得清晰显明的概念，也便于对资料进行全盘叙述。因此，在统计研究时，须将原始标识值区分为适当大小的各组，并显示出组内标识值的频数（个数），列成一表，由任一套标识值，经这样系统的分组而编制成的表，称为一个频数分配。

频数分配分组的原则：编制频数分配时，分组之大小，组数多少虽无具体标准，要根据资料的性质与研究的目的而定，但须遵守以下两个原则：

(1) 要使归入任何一组的标识值都可看作与该组组值相等而不致发生很大的差异。

(2) 为使原标识值编制成一个简单的结构，在不违反(1)的可能范围内，可将组的宽度尽量放大。

根据经验，通常将一套标识值分为10—15组，最多不超过25组为适宜。这样较能符合上述原则。若分组太少，必然引致较大的误差，若组数太多，则制成的频数分配又过于冗长。

编制频数分配的方法和步骤：

1. 计算全距，一套标识值中最小标识值和最大标识值之差，称为全距。以表2-1为例，找出：其中最小标识值是80.1，最大标识值是104.8，两者之差是 $104.8 - 80.1$

表2-1 某市某年3岁女童150人的身长记录(厘米)

80.1	100.1	97.0	96.7	97.9	100.7	86.2	91.7	94.7	90.8
82.5	102.6	99.1	96.6	99.3	85.2	89.2	90.6	95.1	93.6
84.4	104.8	101.3	98.7	101.5	87.1	89.0	92.7	96.8	92.7
87.2	83.5	103.2	101.6	84.4	88.4	91.8	93.6	99.2	94.4
89.3	84.2	82.3	84.5	87.9	89.4	91.9	94.5	86.9	95.6
89.1	86.5	85.0	87.1	89.3	90.4	92.1	95.0	89.3	96.3
91.3	89.7	87.4	89.8	88.7	90.2	92.9	97.2	91.4	90.3
90.5	88.9	88.1	88.2	91.1	93.0	95.6	98.7	90.0	93.5
92.4	90.0	88.0	90.7	91.7	93.8	94.4	87.3	93.9	92.8
92.6	90.0	90.8	90.1	93.2	94.4	97.3	89.0	92.9	94.3
94.1	92.8	90.3	92.8	93.6	94.8	98.3	88.5	94.0	96.0
94.8	92.3	93.3	93.1	95.1	97.0	84.5	91.1	94.3	93.4
97.1	95.3	93.7	95.1	94.9	99.4	86.4	91.7	96.5	92.5
96.2	94.3	94.2	94.6	96.4	100.4	89.1	93.2	98.4	95.5
99.5	93.5	95.1	96.2	99.5	85.7	88.4	92.5	91.1	97.3

= 24.7 厘米。这就是这一套标识值的全距。

2. 确定组距和组数：组距即组的宽度。依据前述分组原则，取全距的 1/10 数值的整数作为组距，可使得出的组数在 10—15 组之间，与分组原则相符。例如表 2-1 的全距为 24.7 厘米，其 1/10 为 2.47，取其整数 2.0 为组距，再以此组距去除全距 $24.7 \div 2.0 = 12.35$ 。于是我们可以估计这一套标识值制成的频数分配，若组距为 2.0，分为 13 组，可以将整套标识值包括在内，也与分组原则相符。因此就确定取 2.0 厘米作为组距。

组距确定后，分组时在开始的一组 and 终末一组不一定用最小最大标识值，但必须分别包括这两个标识值。为了使分组清晰、整齐，便于归纳统计，我们可以从接近于最小标识值的整数开始，按相等的组距依次分组至包括最大标识值的一组止。如制作表 2-1 的频数分配时，可以由 80.0 开始，依次以 2.0 为组距，列成如表 2-2 所示各组，直至包括最大标识值 104.8 的一组作为末一组为止，恰为 13 个组。

3. 列表并标出组限：各组开始的数值及截止的数值称为组限。开始的数值称为下限，截止的数值称为上限。一般情况下，一个组的上限正是下一个组的下限。由于组距即组的宽度，故组距的值等于相邻两组的下限的差或相邻两组上限的差。

组的下限和上限的均数称为组中值或组值，即组的中心值，是代表组的典型数值。现将表 2-1 列成频数分配如表 2-2。

表 2-2 某市某年 3 岁女童 150 人之身长频数分配

身 长 (厘米)	划 记	人 数 (频数)
80.0-	I	1
82.0-	III	3
84.0-	III III	8
86.0-	III III	10
88.0-	III III III	19
90.0-	III III III III	23
92.0-	III III III III I	26
94.0-	III III III III	24
96.0-	III III III II	17
98.0-	III III	10
100.0-	III I	6
102.0-	II	2
104.0-106.0	I	1
总 计		150

如表 2-2 第一组的下限是 80.0，上限是 82.0，（其较精确的数值是 81.99）；第二组的下限是 82.0，上限是 84.0。组距是 $82.0 - 80.0 = 2.0$ 。第一组的组中值是 $(80.0 +$

$82.0) \div 2 = 81.0$ 。

频数分配以直方图表示法：以横座标表示标识值，以纵座标表示标识值的个数，在各组的范围内竖立起高度和各组频数成正比的长方形，构成直方图，各长方形底的中点适位于各组值之位置。这种图示法是以长方形的面积代表频数。图2-1是表2-2的直方图。

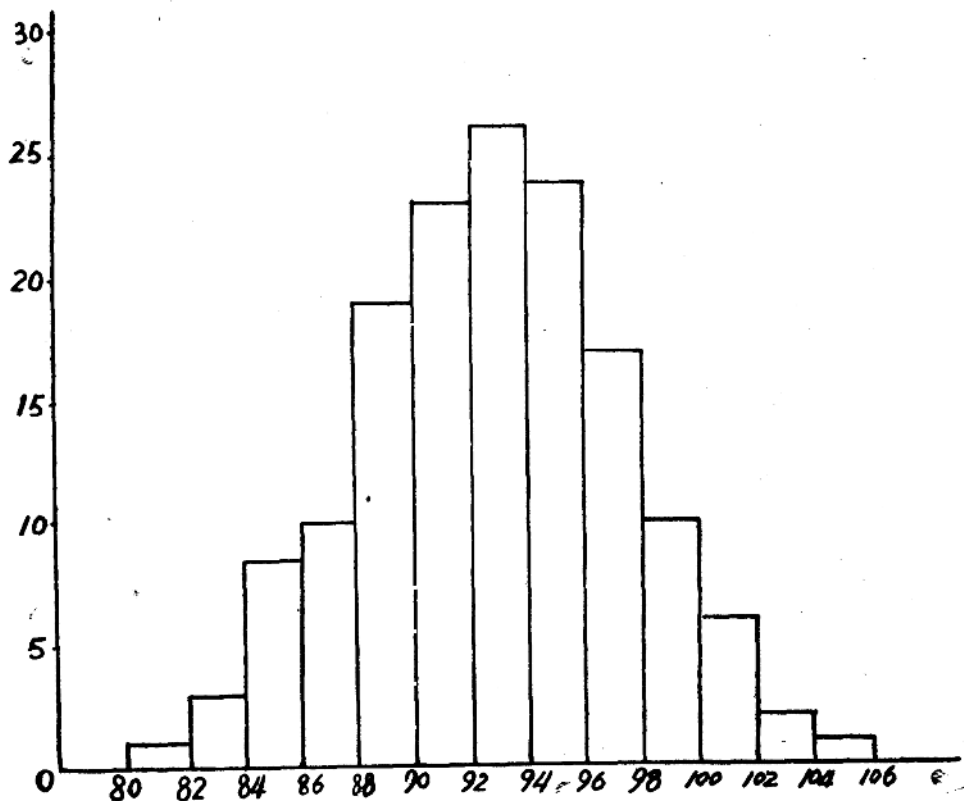


图2-1 某市某年3岁女童150人身长之直方图

由表2-2及图2-1可初步了解表2-1所记录的150个标识值的综合情况，这个初步了解并不是在表2-1原记录中所能获得的，是进行频数分配后才获得的。

第五节 统计 表

资料经分组整理后，若单纯用文字来叙述现象，分析情况，会使阅者感到繁冗而不清晰，且不利于比较和说明问题。为便于对事物现象的分析，使阅者能一目了然，就须制列统计表和必要的图示。

统计表的意義就是将统计资料用表格形式来概括现象的实况，使资料的主要内容能从表的行列中显示出来，免除繁琐的文字叙述。