

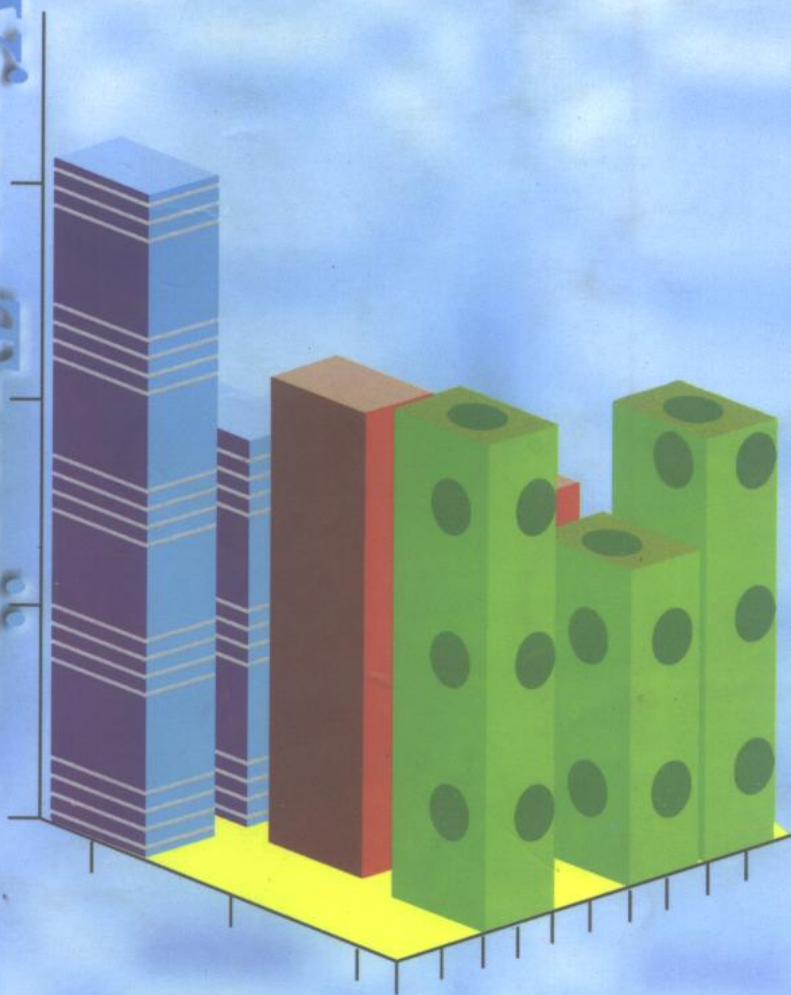
● 成人医学教育教材 ●

医学

统计学

主编 沈 毅 严曰树

Yixue Tongjixue



上海医科大学出版社

成人医学教育教材

医学统计学

主 编 沈 毅 严曰树

副主编 章娜英

编 委 (以姓氏笔画为序)

严曰树 杨廷仕 李跃军

沈 毅 章娜英



上海医科大学出版社

图书在版编目(CIP)数据

医学统计学/沈毅,严曰树编著. —上海:上海医科大学出版社,1999.9
ISBN 7-5627-0521-6

I. 医… II. ①沈… ②严… III. 医学统计 IV. R195.1

中国版本图书馆 CIP 数据核字(99)第 35859 号

责任编辑 沈彬源

责任校对 王秀群

医学统计学

主编 沈毅 严曰树

上海医科大学出版社出版发行

上海市医学院路 138 号

邮政编码 200032

新华书店上海发行所经销

山东省莱芜市印刷厂印刷

开本 787×1092 1/16 印张 10.75 字数 261 000

1999 年 9 月第 1 版 1999 年 9 月第 1 次印刷

印数 1—6 000

ISBN 7-5627-0521-6/R·491

定价:22.00 元

如遇印、装质量问题,请直接与印刷厂联系调换

(地址:山东省莱芜市凤城西大街 149 号 邮编:271100)

内 容 提 要

本书根据成人教育特点,密切结合医学卫生实际,深入浅出地讲述各种资料类型的统计描述、统计推断、统计表和统计图的绘制、医学参考值范围制订方法、实验设计、统计方法在实验室质量控制中的应用,并编写了复习思考题、计算题、函数型计算器的使用方法。

本书可作为各类成人医学教育(含医学职业技术学院),包括临床医学、预防医学、口腔医学、医学检验、全科医生、社区护理等专业的教材,亦可作为医药卫生人员继续教育辅导的基础教材,并可作为各类医药卫生人员特别是检验工作者的参考用书。

EU02 64²⁰

前 言

医学统计学是研究医学中随机现象规律性的一门学科。在医学研究中存在大量的随机现象,因此,医学统计学是医学生必须掌握的一门学科。

编者从事多年的成人医学教育,深感成人医学教育不同于医学本科生教学,尤其在教材的使用上,至今还没有一本适合于成人医学教育的《医学统计学》教材。编者汇总了在原浙江医科大学、浙江中医学院、云南大理医学院成人教育部和浙江省卫生干部进修学校使用多年的成人医学教育的《医学统计学》讲稿,以及在浙江省预防医学会、浙江省护理学会、全国中等卫校师资讲习班、杭州市医药卫生人员继续教育辅导所作《医学统计学》学术讲座的讲稿,并作了修改和补充,编写了这本适合于成人教育的《医学统计学》教材。

本书主要适合于各类成人医学教育(含医学职业技术学院),包括临床医学专业、预防医学专业、口腔医学专业、医学检验专业、全科医生专业、社区护理专业等的学历教育与非学历教育(专业证书班),也可作为医药卫生人员继续教育的基础教材以及各类医学卫生人员和医学统计学教育工作者参考用书。全书共分十三章,其中第十三章专供检验专业使用,标有*号的为选学内容。本书的附录还安排了一些复习思考题,供学员课后复习。

本书十分注意成人教育的特点,密切结合医学实际,深入浅出,通俗易懂,力求充分体现科学性、实用性和先进性。由于我们是初次编写成人教育教材,虽力求提高质量,反复研讨,数易其稿,但限于水平,仍会有不足与错误之处,诚挚希望有关专家及读者,不吝批评指正,以便我们修正,更好地为读者服务。

本书在编写过程中得到了原浙江医科大学成人教育学院、浙江中医学院、浙江省卫生干部进修学校、云南大理医学院成人教育部、杭州市卫生局等单位领导的大力支持,亦得到部分高校教授的指导,浙江省预防医学会常务理事、浙江省卫生学校原卫生学科主任严启之老师对本书的编写起了积极的促进作用,浙江大学医学院卫生统计学教研室倪春春老师对本书中的例题进行了精心的校对,我们谨表示衷心的感谢。

沈 毅 严日树

1999年7月

目 录

第一章 概论	(1)
第一节 随机现象与统计推断	(1)
第二节 统计学的几个基本概念	(1)
第三节 统计工作的基本步骤	(3)
第二章 统计表与统计图	(5)
第一节 统计表	(5)
第二节 统计图	(6)
第三章 计量资料的集中趋势和离散程度	(11)
第一节 数值变量资料的频数分布	(11)
第二节 集中趋势指标	(13)
第三节 离散程度指标	(17)
第四章 正态分布与临床参考值范围	(20)
第一节 正态分布概念	(20)
第二节 标准正态分布	(21)
第三节* 制定临床参考值范围的原则和步骤	(22)
第四节 正态分布资料临床参考值的制定方法	(23)
第五节 偏态分布资料临床参考值的制定方法	(24)
第五章 总体均数的估计和假设检验	(26)
第一节 均数的抽样误差	(26)
第二节 t 分布	(27)
第三节 总体均数可信区间的估计	(28)
第四节 假设检验	(28)
第五节* 正态性检验	(35)
第六节 假设检验应注意事项	(36)
第六章 方差分析	(39)
第一节 完全随机设计资料的方差分析(单因素方差分析)	(39)
第二节 随机区组(配伍组)设计资料的方差分析	(41)

第三节	多个样本均数间的两两比较	(43)
第七章	相对数	(45)
第一节	常用相对数	(45)
第二节	应用相对数应注意的几个问题	(48)
第三节	率的标准化	(49)
第四节	总体率的估计	(50)
第五节	大样本率与总体率比较的 u 检验	(51)
第六节	两个大样本率比较的 u 检验	(51)
第八章	χ^2(卡方)检验	(53)
第一节	四格表资料的 χ^2 检验	(53)
第二节	配对计数资料的 χ^2 检验	(55)
第三节	行 $\times$ 列表资料的 χ^2 检验	(57)
第四节*	四格表资料的确切概率法	(59)
第九章*	二项分布与泊松分布及其应用	(61)
第一节	二项分布	(61)
第二节	小样本率与总体率的比较	(63)
第三节	泊松分布	(64)
第四节	等范围计数资料的总体均数估计	(65)
第五节	等范围计数资料的假设检验	(66)
第十章	秩和检验	(69)
第一节	配对比较的符号秩和检验	(69)
第二节	两样本比较的秩和检验(Wilcoxon 两样本比较法)	(70)
第三节	多个样本比较的秩和检验	(73)
第四节	随机区组设计的多个样本比较的秩和检验	(75)
第十一章	相关与回归	(77)
第一节	直线相关	(77)
第二节	直线回归	(79)
第三节	等级相关	(81)
第四节	作相关回归分析应注意事项	(84)
第十二章	实验设计	(86)
第一节	实验设计的基本要素	(86)
第二节	实验设计的基本原则	(88)
第三节	常用的几种实验设计方法	(91)

第十三章 *	统计方法在实验室质量控制中的应用	(94)
第一节	实验室内部的质量控制	(94)
第二节	实验室间的质量控制	(103)
第三节	质量控制中应用统计方法应注意事项	(108)
附录一	复习思考题.....	(110)
附录二	函数型电子计算器的使用方法	(127)
附录三	常用统计学用表	(133)

第一章 概 论

第一节 随机现象与统计推断

一、随机现象(random phenomenon)

在观察自然和社会现象中,有些是必然发生的,也就是确定发生的现象,例如把石块向上抛后必然落下地的现象;同性电荷必然相斥的现象;人在长期缺氧情况下必然死亡的现象。但是更多遇到的是不确定的现象,例如吸烟的人是否患肺癌的现象;用青霉素治疗大叶性肺炎病人是否能治愈的现象;与病毒性肝炎密切接触的人是否患病毒性肝炎的现象。有时我们把随机现象是否发生称为随机事件是否发生,上述吸烟的人患肺癌事件;用青霉素治疗大叶性肺炎治愈事件;与病毒性肝炎病人密切接触的人患病毒性肝炎事件都是随机事件。

二、统计推断(statistical inference)

随机事件是否发生虽然是不确定的,但应用统计方法,通过对局部的观察可发现其有统计规律性,称为统计推断。例如,在其他条件基本控制均衡的前提下,各国科学家对吸烟的人和吸烟的人作对比观察,都得到同样推断,吸烟者患肺癌的可能性远超过不吸烟者,告诫人们要控制吸烟。又如,对部分密切接触病毒性肝炎病人和不密切接触者作对比观察,可作出密切接触者易得病毒性肝炎的推断,从而提出要作好预防工作;再如,通过青霉素和其他药物对部分大叶性肺炎病人疗效的对照观察,可推断青霉素和其他药物治疗效果有否差别。

由于在医学研究中存在大量的随机现象,因此《医学统计学》是探索医学研究中随机现象及其规律性的一门学科,每一个医学工作者都应掌握这门学科的基本原理和方法,以便更好地做好医疗卫生工作。

第二节 统计学的几个基本概念

一、总体(population)

为研究某地健康成年男子的血红蛋白情况,该地每一个健康成年男子就是一个观察单位,每一个观察单位都可以测得一个血红蛋白值,所有观察单位血红蛋白值的集合称为总体。又如研究某地儿童的蛔虫卵的感染情况,该地每一个儿童就是一个观测单位,每一个观察单位检查蛔虫卵感染结果是定性的,可取阳性值或阴性值,所有观察单位这些值的集合亦形成一个总体。一般来说,总体就是根据研究目的确定的同质研究对象的全体。确切地讲,是同质的所有观察单位某种变量值的集合。如上述该地所有健康成年男子的血红蛋白值就构成一个总体。

二、样本(sample)

在医学研究中,有时不可能也没有必要直接对总体进行研究,所以在实际工作中经常是从总体中随机抽取一部分观察单位进行研究,从总体中随机抽取的一部分观察单位称为样本。所谓随机,就是总体中所有观察单位都有同等可能被抽到。所以要随机抽样,是为了保证样本对总体有代表性。

三、概率(probability)

在一定的条件下某事件发生的可能性的,称为该事件发生的概率。例如,在一个袋中有 99 个红球,1 个白球,在这个袋中,随机摸取 1 个球,则出现红球事件的概率为 $99/100$,出现白球的概率为 $1/100$ 。概率可用 P 表示,显然, $0 \leq P \leq 1$ 。通常,一个事件发生的概率小于或等于 0.05,称为小概率事件,而小概率事件在一次试验中,基本上是不发生的。例如在一个袋中装有 999 个红球,1 个白球,如果在这个袋中随机摸 1 个球,则出现白球的事件基本是不会发生的,因为出现白球的概率为 $P = 0.001$ 。有时我们亦把 $P \leq 0.05$ 或 $P \leq 0.01$ 作为小概率事件。

四、变量(variable)

我们把所研究事件的指标称为变量。例如,研究某地健康成年男子的血红蛋白的情况,血红蛋白就是变量;研究某药对某病治疗的结果,治疗的结果就是变量;变量可分为数值变量和分类变量(定性变量)两种。

(一) 数值变量(numerical variable)

变量值取数值(定量)的变量,称为数值变量。例如研究 100 名健康成年男子的白细胞总数,每一名经定量检验结果,白细胞总数都是取数值的,白细胞总数可认为是数值变量。类似的,如体重、体温、血压等都可以用工具定量地测得其数值,都可视作数值变量,对应得到的变量值亦称为计量资料。

(二) 分类变量(categorical variable)或称定性变量

可分为二项分类变量和多项分类变量。

1. 二项分类变量 例如,60 名儿童蛔虫卵检验结果,可视为变量。变量值可取阳性和阴性两项,蛔虫卵检验结果可认为是二项分类变量,对应得到的变量值的个数亦称为计数资料。例如,60 名儿童蛔虫卵检查结果 36 名阳性,24 名阴性,这种资料是计数资料。某药治疗某病的结果如分为治愈、未愈两项,则对应得到的资料也是计数资料。

2. 多项分类变量 又可分为无序的和有序的两种。

(1) 无序多项分类:例如,成年人血型可视为变量,其变量值可取 A、B、O、AB 型无等级顺序四项,对应得到的资料亦是计数资料。例如,70 名成年人血型人数分类,A 型为 23,B 型为 20,O 型为 24,AB 型为 3。这种资料是计数资料。

(2) 有序多项分类:例如,80 名肾炎病人尿中红细胞检查结果可视为变量,其变量值如取 +、++、+++、++++ 有等级顺序四级,对应得到的资料,称为等级资料。例如,80 名肾炎病人,尿中血红蛋白数 + 为 30,++ 为 32,+++ 为 10,++++ 为 8,这种资料是等级资料。又如,某药治疗某病,治疗结果如分为治愈、显效、好转、无效四个等级顺序的四项,对应得到的资料可视为等级资料。

资料类型可互相转化,例如白细胞总数检查时,如每一个人的白细胞总数视为变量,如变量值取数值,对应得到的资料就可视为计量资料。如变量取正常和不正常两项,对应得到

的资料就可视为计数资料,如变量值取偏低、正常、偏高三项,就可视为等级资料。同样,如把蛔虫卵阴性和阳性,分别计为数值 0 和 1,就可转化为计量资料。

五、误差(error)

统计学常把误差分为抽样误差、随机测量误差、系统误差和过失误差四项。

(一) 抽样误差(sampling error)

由于抽样原因引起的样本指标与总体指标间的差别称为抽样误差。例如在某地普查结果,其恶性肿瘤的发病率为 512/100 000,如抽样调查,其发病率就不一定等于 512/100 000。又如,普查某厂成年男职工的红细胞平均数为 $4.8 \times 10^{12}/L$,如果抽样调查,其红细胞平均数就不一定等于 $4.8 \times 10^{12}/L$ 。

(二) 随机测量误差(random measure error)

例如,把一个水样标本,连续测定数次,其测量结果不一定相同,这种误差就是随机测量误差。

(三) 系统误差(system error)

由于人为原因引起的误差称为系统误差。例如,由于仪器不准,天平未校正,试剂不标准等,都可引起系统误差。

(四) 过失误差 由于操作人员操作上的失误引起的误差称为过失误差。

抽样误差和随机测量误差是不能避免的,但可控制。系统误差和过失误差如能把产生的原因消除就可以避免。

第三节 统计工作的基本步骤

统计工作的基本步骤可分为以下四步:

一、设计(design)

设计是整个统计工作的关键,也就是整个统计工作全过程的设想和安排,例如:研究的目标和假设是什么?如何确定观察对象和观察指标?样本大小如何确定?怎样控制误差等等。一个好的设计往往可达到事半功倍的效果(详见第十二章)。

二、搜集资料(collection data)

按资料性质分,可分为日常性资料、专题调查资料和实验研究资料。搜集资料做到准确、完整、可靠、及时。

日常性资料包括日常医疗工作记录,如门诊、住院病历;专门报告卡,如出生、死亡、传染病报告卡;统计报表,如疫情月报表、医院年报表等。

专题调查研究是对研究对象不进行任何干预,客观反映实际情况,如研究某地健康成年男子的血红蛋白水平;研究某地小学生的蛔虫卵感染情况。

实验研究是研究者根据研究目的,人为地进行一些干预措施,例如,为观察局部温热疗法治疗生癌小鼠的疗效,先人为地给小鼠注射致癌物质,使其患癌(详见第十二章)。

三、整理资料(sorting data)

从搜集资料得到的原始资料往往是零乱的、分散的,需要进行整理,使其系统化、条理化,以便于统计分析。

(一) 检查核对

观察资料项目填写是否完整正确、有否遗漏,如一般项目中有是否漏填了姓名、性别;例如3岁女童身高(cm)记录数据中有928,就需检查是否遗漏了小数点,在核对时发现问题应及时补正。

(二)设计分组

1. 数量分组 如按年龄、身高值、血红蛋白值、血压值、胆固醇值等分组。
2. 质量分组 如按职业、性别、血型、疾病种类等分组。

(三)拟订整理表

根据研究目的拟订合适的整理表,如需了解某地健康成年男子血红蛋白的分布特征,可按血红蛋白值分组,了解小学学生蛔虫卵感染情况,可按性别分组,拟订整理表(表1.1和表1.2)。

把各变量值划分归属在各个组,如将上述血红蛋白值和蛔虫卵阳性原始变量值分别划计在各个组段,可得各组的频数(人数)。

表 1.1 某地 108 名成年男子血红蛋白(g/L)

血红蛋白量	划 记	频 数
105 ~		
110 ~		
115 ~		
120 ~		
.....		
.....		
170 ~		

表 1.2 某小学 100 名小学生蛔虫感染情况

性 别	阳性划记	频 数
男		
女		

四、分析资料(analysis of data)

资料分析的目的是计算有关统计指标,反映事物的特征,阐明事物内在的规律性。

(一)分析事物特征

搜集得到的原始资料经整理后比较条理化、系统化,可根据研究目的,应用统计方法揭示分布特征,进行比较分析。如根据所得到的资料,计算其集中趋势指标,如平均数;离散程度指标,如标准差;计算相对数指标,如率和构成比;根据样本指标推断总体指标所在的范围,例如,通过对某地3岁女童身高的抽样调查,可以推断该地3岁女孩身高总体平均水平的一个范围;再如,通过某种中药和某种西药治疗乙脑效果的样本比较信息,推断该中药与西药治疗乙脑效果是否有差别。

(二)分析事物内部构成

通过对事物内部构成的情况调查,分析其构成是否合理,例如,通过对各科住院病人的构成的调查,分析各科病床设置是否合理;通过男、女性别比的调查,分析男、女性别构成是否失调。

(三)分析事物间相互关系

客观事物之间往往存在着内在的联系,例如通过调查研究,可研究吸烟与慢性支气管炎的关系。

第二章 统计表与统计图

第一节 统计表

统计表是表达统计资料及其指标的一种表格,这里所指的表,是指统计分析阶段的统计分析表。

一、绘制统计表的通则

1. 标题写在表的正中、上方,简明扼要地说明表的主题,必要时写上时间和地点。
2. 一般把被分析的事物置在横标目,指标置在纵标目。
3. 主要横线最多只能有四条,除顶线、底线、纵标目下边的线及合计上边的横线外,一般都可以省去,在总标目栏,有时可有短线。在表中,如只需个别列出合计数字,可在相应列的合计栏上方画一短线。
4. 数字一律用阿拉伯数字,位数应对齐,同一列数字小数位数保持一致。
5. 如有备注,可用*标出,并在表下注明。

二、统计表的修改

表 2.1 所示统计表的缺点是:线条不符合要求;项目中调查人数及沙眼人数重复。

表 2.1 某乡四个村儿童沙眼感染率(%)

情况 村别	项 目	例 数	感染率(%)
甲村	调查人数	84	82.14
	沙眼人数	69	
乙村	调查人数	94	77.66
	沙眼人数	73	
丙村	调查人数	74	68.92
	沙眼人数	51	
丁村	调查人数	62	54.84
	沙眼人数	34	
合计	调查人数	314	72.29
	沙眼人数	227	

可修改成如表 2.2。

表 2.2 某乡四个村儿童沙眼感染率(%)

村 别	调查人数	沙眼人数	感染率(%)
甲村	84	69	82.14
乙村	94	73	77.66
丙村	74	51	68.92
丁村	62	34	54.84
合计	314	227	72.29

第二节 统计图

统计图是把统计资料用图形的形式表达出来,它用点的位置,直条的长短,或面积的大小把资料反映的趋势、分布和事物与事物之间的数量关系形象地表达出来。

一、绘制统计图的通则

1. 应根据资料性质选择适合的图形。
2. 标题写在图的正中、下方,简明扼要地说明主题,必要时写上时间和地点。
3. 如图形有纵、横轴,需要注明指标、单位,标明尺度,纵轴尺度自下而上,横轴尺度自左至右,一律有由小至大。
4. 必要时用图例注明。
5. 除构成图外,图形纵横坐标之比一般以 5:7 左右为宜。

二、常用统计图

(一) 直条图

分单式直条图和复式直条图,适用互相独立的资料,由表 2.2 资料,可绘制单式直条图,见图 2.1。

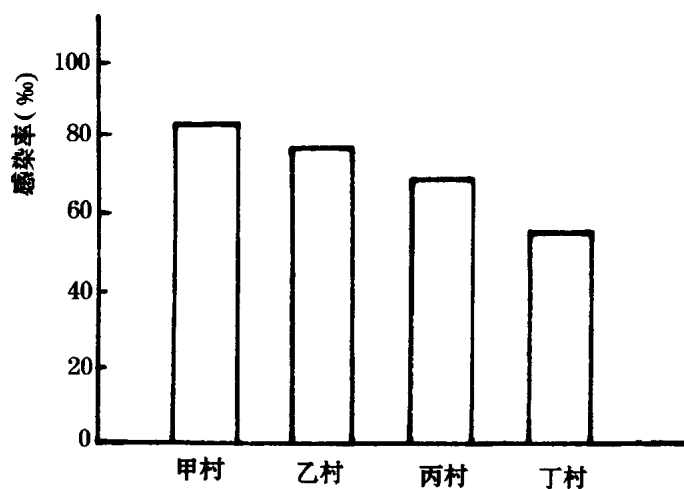


图 2.1 某乡四个村儿童沙眼感染率

由表 2.3 的资料,可绘制复式直条图,见图 2.2。

表 2.3 某县四种恶性肿瘤男、女的死亡率(1/100 000)

肿瘤类型	男	女
食管癌	28	10
胃癌	23	12
肝癌	11	6
肺癌	5	3

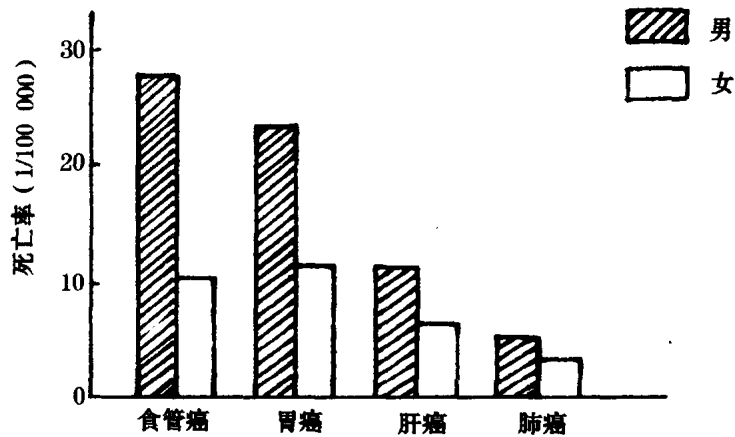


图 2.2 某县四种恶性肿瘤性别死亡率

(二) 构成图

用于主要表示构成比的资料,构成图包括百分直条图和圆形图,如表 2.4 资料可画成下列构成图。

表 2.4 急性白血病患者血型分布

血 型	例 数	构成比(%)
O 型	118	32.1
A 型	116	31.5
B 型	98	26.6
AB 型	36	9.8
合计	368	100.0

1. 百分直条图 先画一线段,总长度作为 100%,画一条长度与这条线段的总长度相等的直条,直条的总面积亦作为 100%,直条内的某一部分直条的长度等于总长度的百分之几,对应部分的直条面积亦是总直条面积的百分之几,见图 2.3。

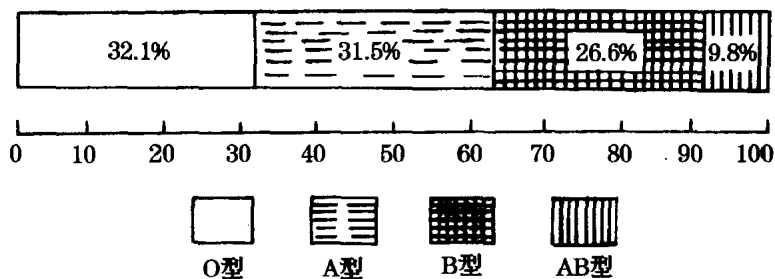


图 2.3 急性白血病患者血型分布

2. 圆形图 以圆的面积作为 100%,对应的圆内扇形面积是圆面积的百分之几,就可表示对应的构成。当圆心角为 360° 时,对应的圆面积是 100%,故圆心角是 3.6° 时的扇形面积就对应圆的面积的 1%,把 3.6° 乘以构成比的分子,就是对应的扇形面积的圆心角的度数。例如,要表示构成比 32.1%,则对应扇形的圆心角就是 $3.6^\circ \times 32.1 = 115.56^\circ$,表 2.4 的资料,亦可以用圆形图表示,见图 2.4。

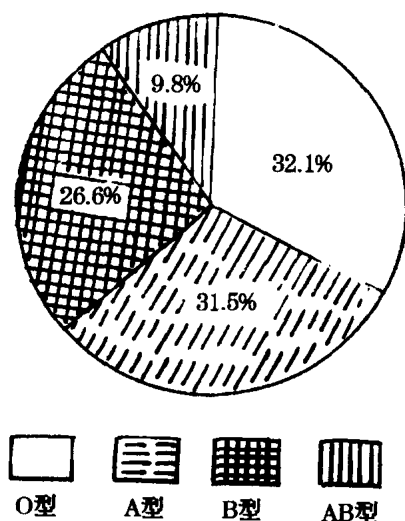


图 2.4 急性白血病患者血型分布

(三) 线图适用于连续性但不表示频数分布的资料,可分为普通线图和半对数线图

1. 普通线图 横轴和纵轴都是算术尺度的线图称为普通线图。所谓算术尺度,就是相等的长度,表示相等的数量。普通线图适用于纵坐标变动幅度不大的资料,表 2.5 资料,可用普通线图表示,如图 2.5。

表 2.5 某地 1949~1957 年婴儿死亡率

年 份	婴儿死亡率(%)
1949	120
1950	95
1951	82
1952	60
1953	50
1954	42
1955	30
1956	20
1957	10

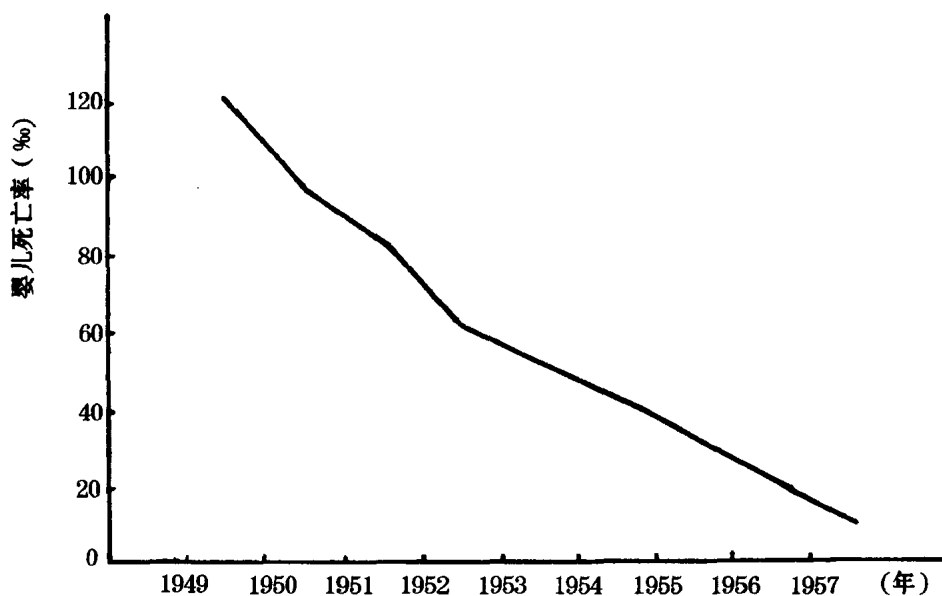


图 2.5 某地 1949~1957 年婴儿死亡率

2. 半对数线图 横轴用算术尺度,纵轴用对数尺度的线图,称为半对数线图。表 2.6 的资料由于伤寒、副伤寒死亡率变化较大,可把死亡率的分子取对数后作为纵坐标,亦可直接把死亡率分子作为纵坐标,见图 2.6。

表 2.6 某地 1965 ~ 1995 年伤寒、副伤寒死亡率

年 份	伤寒、副伤寒死亡率(1/100 000)	对数值(1/100 000)
1965	18.0	1.255 3
1970	9.2	0.963 8
1975	5.0	0.699 0
1980	3.8	0.579 8
1985	1.6	0.204 1
1990	0.8	-0.096 9
1995	0.3	-0.522 9

图 2.6(1)的纵轴是伤寒、副伤寒的死亡率分子的对数值,纵轴对 lgY 来说仍是算术尺度;图 2.6(2)的纵轴是伤寒、副伤寒死亡率,纵轴 Y 是对数尺度。

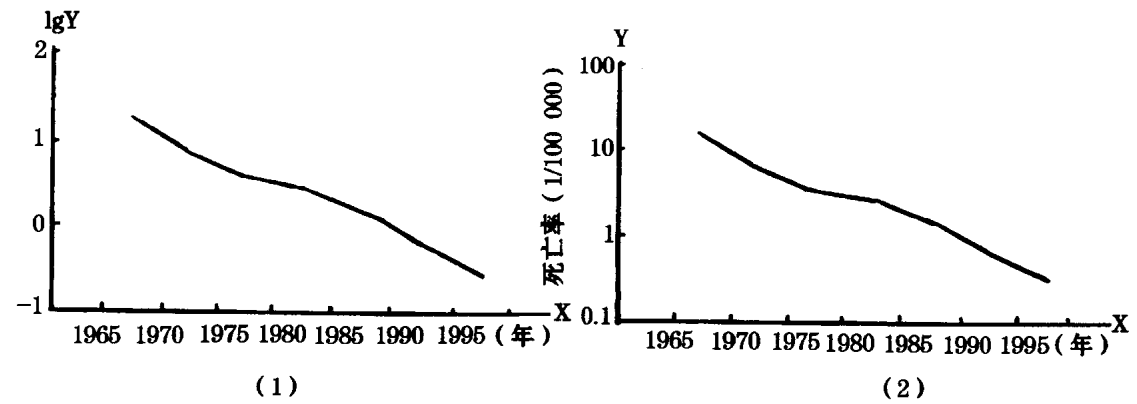


图 2.6 某地 1965 ~ 1995 年伤寒、副伤寒死亡率变化

(四)直方图

适用于数值变量的频数表资料,如表 2.7 资料可画成直方图,见图 2.7。

表 2.7 某市 120 名 12 岁男孩身高(cm)的频数分布

身高组段	频 数
125 ~	1
129 ~	4
133 ~	9
137 ~	28
141 ~	35
145 ~	27
149 ~	11
153 ~	4
157 ~	1