

陆守曾 主编

医学统计学

Medical Statistics



中国统计出版社
China Statistics Press

医学统计学

主 编 陆守曾
副 主 编 陈 峰 蔡 辉
编 者 陆守曾 陈 峰 蔡 辉
顾海雁 陈佩珍 黄水平
学术秘书 荀鹏程

 中国统计出版社
China Statistics Press

(京)新登字 041 号

图书在版编目 (CIP) 数据

医学统计学/陆守曾主编.
-北京:中国统计出版社 2002.1
ISBN 7-5037-3675-5

I. 医…
II. 陆…
III. 医学统计
IV. R195.1

中国版本图书馆 CIP 数据核字(2001)第 076081 号

医学统计学

作 者/陆守曾主编
责任编辑/吕 军
封面设计/张建民
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街 75 号 邮政编码/100826
办公地址/北京市丰台区西三环南路甲 6 号
电 话/(010)63459084、63266600—22500 (发行部)
印 刷/科伦克三莱印务(北京)有限公司
经 销/新华书店
开 本/787×1092mm 1/16
字 数/600 千字
印 张/25.125
印 数/1—6000 册
版 别/2002 年 1 月第 1 版
版 次/2002 年 1 月北京第 1 次印刷
书 号/ISBN 7-5037-3675-5/R·2
定 价/48.00 元

版权所有。未经许可,本书的任何部分不准以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。
中国统计版图书,如有印装错误,本社发行部负责调换。

序

一切真知都来源于实践，尤其是医学这门实践性极强的学科。但是，由于生物体之间存在着个体差异，人类的机体反应更是既受到自然环境的影响，又受到种种社会因素的制约，因此其表现千变万化，难以捉摸，这就给人们总结实践经验带来不少困难。

数理统计学渗入到医学领域为克服上述困难提供了有力的武器。既为医学的发展服务，从而发挥强大的作用；又在服务过程中遇到了不少医学领域特殊形式和特殊需求的问题。在解决和处理这类问题的过程中，又充实、拓宽和发展了这门学科。于是就逐步形成了医学统计学这一边缘性的新学科。

半个多世纪以来，随着医学科学的迅速发展，医学统计学也日益发展成熟。近年来，随着信息技术的飞速发展，计算机的普遍应用，各种功能优异的软件包的问世，使许多用途广泛、威力强大但计算量惊人的统计分析方法纷纷引入本学科领域，这门学科的重要性也越来越为医学界所公认。这就是近年来各高校都将医学统计学列为医学各专业学生必修课的重要原因，尤其在研究生的培养中，更将本学科的学习作为培养基本科研技能的必要内容。

本书主编毕生从事医学统计学的教学与研究，积累了极其丰富的实践经验，是本学科的知名学者。本书是在其许多年来所编写有关专著基础上不断修改，补充和更新而成，内容有深度，有广度，但又深而不奥，广而不杂，且理论密切联系实际，同时结合大量医学实例，阐述基本原理及其应用方法，文字简练，文笔流畅，编排富有新意，在共性中体现个性，各章内容既互有联系，又独立成篇，便于参阅。

本书既可作为医学各专业本科生的教材，也可用于研究生的教学，对于从事医学研究的科研、教学人员也是一本有益的参考书。

针对不同需要，全书内容分为基础章节、基本章节和参考章节，为读者提供了方便。

相信本书的问世必将受到广大读者的欢迎。乐为之序！

史秉璋

于上海第二医科大学

2001年10月

前 言

医学统计学是医学科学的一个组成部分，是医学院校各专业本科生和研究生的公共必修课。本学科与医学其它学科之间的联系具有非同一般的广度和深度，尤其在医学研究设计、数据分析和结果评价方面更为明显。

自从我的导师郭祖超教授编著的《医学与生物统计方法》一书于 1948 年出版，首次将医学统计学系统地介绍到国内，至今已有半个多世纪。半个多世纪以来，随着医学研究领域的不断拓展和研究内容的不断深入，医学统计学理论亦得到了长足的发展，统计方法层出不穷，应用面亦越来越宽广；计算机和软件技术的飞速发展，使医学统计学的应用变得越来越易于实现，研究者已经从繁琐的数字运算中解脱出来，统计学应用成为一种乐趣，大型资料的分析成为可能。医学统计学的教学也发生了深刻的变化：学生基础学科水平的提高，尤其是数学水平的提高，使教学的起点明显地提高了；教学重点发生了转移，教师可以将主要精力放在基本概念、基本思想、研究设计技巧的讲解以及计算结果的解释上，无需在黑板上再作繁琐的数值演算。医学统计学不再是数据的罗列和公式的堆砌，而是研究设计的艺术和信息表达的艺术。

南通医学院开设的《医学统计学》是江苏省一类优秀课程，这是我院老、中、青三代医学统计学教师长期不懈努力的结果。本书融入了我们多年积累的教学经验和心得，在编撰中力图反映自己的特色，并依据下列编撰原则：

(1) 以医学研究中实际问题为本，采用数理统计方法以及其他方法为手段，以分析有关数据资料，并归结于对医学研究结果的推断。

(2) 注重基本概念、基本思想的阐述，并用以指导对各种方法的理解、选择、评价和对分析结果的解释。

(3) 注重医学统计学思维的培养，始终贯穿“医学实际问题—研究指标—变异和分布—抽样误差—统计推断—对医学实际问题的解释”这一主线。

(4) 强调研究设计在医学研究中的地位和作用，并作了较为全面、系统、深入的阐述。

(5) 尽量适应本科生、研究生的可接受性；尽量满足实际研究工作的需要。

本书共 22 章，第一、二、三、四、五共 5 章，主要为基本概念，包括研究指标的分类、变异、分布、抽样误差、参数估计和假设检验等；第六、七、八、九共 4 章，为定量资料、定性资料、等级资料单指标的分析；第十、十一、十二共 3 章，为两指标的相关与回归分析，包括三类资料的相关、定量资料的

直线回归和曲线回归；第十三、十四、十五、十六共4章，为研究设计，包括研究设计总论、实验设计、临床试验设计和调查设计；十七、十八两章为多元统计分析简介，包括多元线性回归、logistic回归、Cox回归、聚类与判别分析；其他分析方法共4章：第十九、二十、二十一、二十二章，包括圆形分布资料的统计分析、综合分析与综合评价、生命统计、统计图表。

本书内容分三个层次：基础章节、基本章节和参考章节。基础章节为本科生必学；基本章节为研究生必学；参考章节供科研工作者参考。各章节分层参见附录D。因此，本书适用于不同学时的《医学统计学》的教学，同时又有参考价值。

本书的编写得到了南通医学院各级领导的大力支持。荀鹏程同志担任本书学术秘书，承担了繁重的文字排版、制图和校对任务，并与赵杨同志共同对书中计算部分进行了复核。朱湘竹、沈毅、朱晓蓉、丁红同志帮助打印和校对了部分书稿。以上诸位同仁通过对本书初稿的研读、推敲和质疑，提出了大量的、深入的、精辟的见解和建议，其中绝大部分已被采纳。谨致以衷心感谢。

陆守曾 谨识于南通医学院

2001年10月

目 录

第一章 绪言	1
1.1 医学统计学	1
1.2 医学统计学的概括内容	1
1.3 几个基本概念	2
1.4 培养医学统计学思维	4
第二章 个体变异与变量分布	6
2.1 个体变异	6
2.2 频数分布	6
2.3 定量资料的统计指标	9
2.4 定性资料与等级资料的统计指标	13
2.5 正态分布	17
2.6 二项分布	20
2.7 Poisson 分布	23
2.8 参考值范围的确定	25
第三章 抽样误差	30
3.1 抽样误差的概念	30
3.2 抽样误差产生的条件	30
3.3 均数的抽样误差及标准误	30
3.4 t 分布	34
3.5 χ^2 分布	35
3.6 F 分布	36
第四章 可信区间	38
4.1 可信区间的概念	38
4.2 均数的可信区间	39
4.3 率的可信区间	41
4.4 事件数的可信区间	43
4.5 方差的可信区间	44
4.6 可信区间的正确应用	44

第五章 假设检验	47
5.1 假设检验的意义	47
5.2 假设检验的思路	48
5.3 假设检验的步骤	50
5.4 假设检验的正确应用	51
5.5 假设检验的几个相关问题	53
第六章 定量资料的分析	58
6.1 样本均数与总体均数的比较	58
6.2 两样本均数比较的 t 检验	61
6.3 t 检验的正确应用	63
6.4 多个均数的比较	64
6.5 方差齐性检验	70
6.6 方差分析的正确应用	72
第七章 定性资料的分析	74
7.1 样本率与总体率的比较	74
7.2 两样本率的比较	74
7.3 多个率的比较	78
7.4 构成比的比较	80
7.5 配对设计两样本率的比较	81
7.6 似然比检验	82
7.7 确切概率法	83
7.8 两事件数的比较	85
7.9 定性资料假设检验的正确应用	86
第八章 等级资料的分析	88
8.1 秩次与秩和	88
8.2 两组比较的秩和检验	89
8.3 多组比较的秩和检验	91
8.4 配对设计资料的秩和检验	93
8.5 配伍组设计资料的秩和检验	95
8.6 秩和检验的正确应用	97
第九章 资料分布的拟合优度检验	99
9.1 Pearson χ^2 检验	99

9.2 Kolmogorov 检验	104
9.3 正态分布的矩法检验	106
9.4 分布拟合优度检验的正确应用	107
第十章 两指标间的直线相关	109
10.1 相关关系与确定性关系	109
10.2 定量资料的相关	109
10.3 定性资料的相关	115
10.4 等级资料的相关	117
10.5 相关分析的正确应用	119
第十一章 两指标间的直线回归	121
11.1 概念	121
11.2 直线回归方程的建立	122
11.3 回归系数和回归方程的意义及性质	123
11.4 回归系数的假设检验	124
11.5 应变量总变异的分解	125
11.6 回归问题的方差分析	126
11.7 直线回归的区间估计	127
11.8 两条回归直线的比较	129
11.9 过定点的直线回归	131
11.10 直线回归与直线相关的区别及联系	133
11.11 回归分析的正确应用	134
第十二章 两指标间的曲线回归	136
12.1 引言	136
12.2 拟合回归曲线的用途	136
12.3 拟合回归曲线的步骤	137
12.4 拟合指数曲线	138
12.5 拟合双曲线	142
12.6 拟合简单抛物线	145
12.7 拟合 logistic 曲线	149
12.8 拟合回归曲线的正确应用	153
第十三章 研究设计(一) —— 总论	155
13.1 研究设计的作用	155

13.2 研究设计的形式	155
13.3 研究因素与混杂因素	158
13.4 对照组及均衡性	159
13.5 指标及其选择	162
13.6 研究对象的随机化	165
13.7 样本含量及其估计	167
第十四章 研究设计(二) —— 实验设计	171
14.1 选择实验设计方法的依据	171
14.2 完全随机设计	172
14.3 配对设计	173
14.4 配伍组设计	177
14.5 交叉设计	179
14.6 拉丁方设计	182
14.7 析因设计	184
14.8 正交试验设计	186
第十五章 研究设计(三) —— 临床试验设计	193
15.1 临床试验的特点	193
15.2 新药的临床试验	193
15.3 临床诊断试验与评价	204
15.4 临床随访研究及分析	212
第十六章 研究设计(四) —— 调查设计	219
16.1 现场数据的信度与效度分析	219
16.2 横断面调查	222
16.3 回顾性调查	225
16.4 追踪性调查	229
16.5 典型调查	233
第十七章 多元回归分析	234
17.1 多元线性回归	234
17.2 logistic 回归	243
17.3 Cox 比例风险回归	247
17.4 逐步回归	249
17.5 多元回归分析的正确应用	252

第十八章 聚类分析与判别分析	254
18.1 距离与相似系数	254
18.2 聚类分析	256
18.3 判别分析	260
第十九章 圆形分布资料的分析	264
19.1 圆形分布资料	264
19.2 圆形分布资料的统计描述	264
19.3 平均角的假设检验	266
19.4 两样本角均数的比较	267
19.5 圆形统计方法的正确应用	268
第二十章 综合分析与综合评价	269
20.1 多个同类研究结果的 Meta 分析	269
20.2 资料的综合分析	277
20.3 综合评价	280
第二十一章 生命统计	284
21.1 生命统计的内容与意义	284
21.2 人口数量与人口构成	284
21.3 出生统计	287
21.4 死亡统计	292
21.5 寿命表分析	296
21.6 疾病统计	307
第二十二章 统计表与统计图	317
22.1 统计表	317
22.2 统计图	319
附录 A 统计学工具表	325
附录 B 练习题	355
附录 C 英汉医学统计学词汇	379
附录 D 章节分层	388
主要参考书目	389

第一章 绪言

1.1 医学统计学

在医学科学研究中，经常通过数字资料表达其过程和结果。医学统计学(*medical statistics*)就是运用概率论和数理统计等数学的原理、方法，研究医学资料的搜集、整理、分析和推断的一门学科。医学的研究对象主要是人体以及与人体的健康和疾病相关的各种因素。由于影响人体的因素错综复杂，而人体对影响因素的反应又往往各不相同，即个体变异普遍存在，必须运用统计方法透过具有偶然性的现象来探测其规律性。

医学统计学与生物统计学(*biostatistics*)、卫生统计学(*health statistics*)是统计学原理和方法在互有联系的不同学科领域的应用，三者间既有区别，又有交叉，故难以截然划定界限。生物统计学应用于生物学研究，从生物范畴的角度来看，显然比医学统计学的范围更广，其原理和方法一般均可应用于医学研究。医学统计学和卫生统计学均应用于医学研究，而前者侧重于医学的生物性方面，后者侧重于公共卫生学的社会性方面。

从上个世纪二十年代起，尤其是四十年代以后，医学统计学才逐渐形成为一门学科。几十年来，随着医学科研工作的发展，医学统计学已成为医学领域中的重要组成部分，并被定为医学本科生和研究生的必修课程。

1.2 医学统计学的概括内容

1.2.1 研究设计

研究设计主要包括实验设计、临床试验设计和调查设计。实验设计通常指实验因素可严格控制的研究设计，其特点为必须对实验对象进行干预，并主要通过实验收集资料，其研究过程多在实验室进行。临床研究的原理本属实验性研究，但由于临床条件不易控制，影响因素复杂，应用较广，且具有特殊要求，因此有人又将临床研究独立出来，形成临床试验(*clinical trial*)设计。调查设计是针对现场调查所进行的研究设计，这类研究的特点是对调查对象一般不进行人为干预，同时以观察和调查的方式搜集资料，研究者只能进行“被动”的观察。有的大型研究设计，常将实验设计与调查设计相结合，如在现场调查研究中对被观察对象进行人为干预，称现场试验，二者相互补充，以完善整个研究设计。

1.2.2 资料分析

(1) 统计指标

简称指标。在特定范畴内，对某特征作出总的描述或平均描述所用之单个数值或成组数值，这是指标的一般涵义。特定范畴可以是总体，也可以是样本。指标可以从范畴内的观察对象直接计算出来，也可以间接计算出来。指标大体上可以归纳为7类：

① 平均值。如血红蛋白平均值。

② 变异值。如血红蛋白标准差、变异系数、方差。

③ 绝对值。如我国艾滋病人数。

④ 相对值。如某阳性率、男女性别构成比。

⑤ 相关值。如某两指标的相关系数、回归系数。

⑥ 区间值。如总胆固醇(TC)的临界高水平范围是 5.20~6.24 毫摩尔/升,即高于 6.24 毫摩尔/升为高水平,低于 5.20 毫摩尔/升为可接受水平(依据 1993 年美国全国胆固醇教育方案)。这个界限及其区间显然与是否需进行治疗以及如何治疗密切相关。

⑦ 综合值。如综合评价值。

有些指标是从以上各类指标派生出来的。如 LD_{50} (半数致死量)是在回归条件下的中位数,属平均值类; RR (相对危险度)是一种比值,属相对值。

(2) 指标的抽样误差及其统计推断

当指标由样本得出时,为样本指标。样本指标与总体指标之间的差异称为抽样误差。基于抽样误差进行的分析称为统计推断,包括区间估计和假设检验两类。

① 区间估计。通常指以一定的概率(正确率或信度),利用样本数据所含的信息,取某个区间对总体指标进行估计。例如,用样本均数对相应的总体均数进行估计,其区间为总体均数的可信区间。

② 假设检验。在医学研究中,假设检验有两种目的。一种是检验差异,其意义在于能否把 A 组的指标与 B 组的指标区分开来,即判断 A 组与 B 组的差异有没有统计学意义,这就是一般医学统计学教科书中所介绍的假设检验方法。另一种是检验优度,其意义在于能否认为样本 A 属于某特定总体,即判断以该总体解释样本 A 有没有统计学意义,习惯上称此为“优度”——样本 A 获自该总体的优度或拟合优度(goodness of fit)。详见第五章和第九章。

(3) 数据分布

对于一组变量值,若以该变量为横轴,数据出现的频数(或频率)为纵轴作图,该数据在坐标系中呈一定的图形,称为数据的分布。每一个变量都有其特定的分布模式。数据分布是统计方法产生的基础,因而选择和应用统计方法时,数据的分布是主要依据之一。常用的数据分布有正态分布、对数正态分布、二项分布、Poisson 分布等。

1.3 几个基本概念

1.3.1 三个步骤

研究设计、资料分析和结论,这是在科研实施的全过程中运用统计学原理和方法依次操作的三个步骤。

(1) 研究设计:这是在医学研究中运用医学统计学的起点,也是高质量地完成整个研究的重要基础。

(2) 资料分析:是在研究设计基础上,通过实验(试验)或调查,将所得数据进行统计学处理的过程。

(3) 结论:是在数据分析的基础上,应用统计学处理的结果,进行统计学推断;同时,依据相应的专业知识,作出专业性的结论。

1.3.2 三类资料

这里的资料指观察单位的原始记录，或从原始记录过录者。根据资料的性质，并参照研究目的，可将其分为三类：定量资料、定性资料和等级资料。

(1) 定量资料(quantitative data)：以定量值表达每个观察单位的某项观察指标，如血脂、心率等。这类资料的特点是：① 各观察单位间只有量的差别；② 数据间有连续性。

(2) 定性资料(qualitative data)：以定性方式表达每个观察单位的某项观察指标，如血型、性别等。这类资料的特点是：① 各观察单位间或者相同，或者存在质的差别；② 有质的差别者之间无连续性。

(3) 等级资料(ranked data)：以等级表达每个观察单位的某项观察指标，如疗效分级、血粘度、心功能分级等。这类资料的特点是：① 各观察单位间或者相同，或者存在质的差别；② 各等级间只有顺序，而无数值大小，故等级之间不可度量。

值得注意的是，由于研究目的及记录方法的不同，各类资料间有时可相互转换，如血压值一般为定量资料，但若将其记为高血压、正常血压和低血压，就成为等级资料。

1.3.3 总体和样本

所谓总体(population)，是指按研究目的所确定的同质研究对象某项观察指标的全体，即全体观察单位。总体指标称为参数(parameter)。例如，在一项研究中，某医生的目的是想观察促红细胞生成素对再生障碍性贫血的治疗效果，则所有再生障碍性贫血病人组成研究总体。总体可分为无限总体和有限总体，前者指总体内的个体数是无限的，如上例的再生障碍性贫血病人，理论上不但包括现在的病人，而且还包括未来的病人；后者指总体内的个体数是有限的，如血常规检验时只取一滴血，其总体显然就是检验对象体内的全部外周血液，是有限的。样本(sample)是指从研究总体中随机抽取具有代表性的部分观察单位，样本指标称为统计量(statistics)。如在促红细胞生成素疗效的研究中，该医生往往是从此类病人中随机抽出部分(如20例)病人进行研究，并以其样本试验结果通过统计分析来推断针对总体的相应结论。样本含量(sample size)是指样本中的观察单位数，常用 n 表示。

同一总体中的个体具有同质性(homogeneity)。同质性是指观察单位具有相同的性质，与之相反的是异质性或间杂性(heterogeneity)。观察单位间的同质性是构成总体的必备条件，也是进行研究的前提，缺乏同质性的观察单位是不能混在一起进行分析的。如不同年龄组男童的身高不能计算平均数，因为所得结果没有意义。当研究的内容不同，对观察对象的全同性要求也不同，即全同性是相对的。例如，在某新药的临床试验中，计算有效率的观察病例必须患同一疾病，甚至具有相同的病型、病情、病程等，对全同性的要求是很严格的；而计算不良反应率，通常可将不同病种的病例合起来统计，此时对全同性的要求只有一条：按规定服用该新药。

1.3.4 概率

概率(probability)是描述事件发生可能性大小的一个度量，常用 P 来表示，取值范围为 $0 \leq P \leq 1$ 。当某事件发生的概率小于或等于0.05时，统计学上称该事件为小概率事件，其涵义为该事件发生的可能性很小，进而认为其在一次抽样中不可能发生，此即为小概率原理。小概率原理是进行统计推断的依据。

1.4 培养医学统计学思维

医学统计学广泛应用数理统计学的理论和方法,这是毋庸置疑的,但这种应用必须以医学理论及其研究内容为载体。基于此认识,编者常以如下逻辑程序概括本学科的内容:从医学实际出发——运用统计理论和方法分析——阐述某个医学实际问题。学习医学统计学不仅在于获取知识,更重要的是培养医学统计学思维。

医学统计学不同于医学的其它学科。医学课程大致可分为形态、功能、临床三大类。医学生经过几年的学习,往往习惯于观察、记忆、判断、操作,而淡化了数量化的思维。如对于各种检查、化验的结果,若只重视“正常”与“不正常”的区分,这是不够的。须知,凡属计量指标,都有其分布型,人们按照不同的分界可以规定出不同的区间,并依此辨别正常与不正常。可见,同属正常值者也有高、低之分,同属不正常者又有轻、重之别;而且正常者与不正常者的分布几乎无例外地都有不同程度的重叠。所以,从总体来看,依照任何指标所作的诊断,都不可避免出现误诊或漏诊。只有建立起清晰的数量化观念,才能在各种分布中既作出分析判断,同时又看到失误的可能性。这是医学统计学的观念。

医学统计学亦不同于数学。医学统计资料中的每一个数据均有其特定涵义,这与抽象的数字不同。尽管数字本身可能是相似的,甚至是重复的,而实际医学资料所反映的生物现象却从来不会完全相同地重复。数学讲推导,推导出来,被证明了就是真理;医学的基础是个体及其反映的观察指标,各人各样,绝无大同,对同样的病作同样的治疗,有些病人有效,有些病人无效,这也是真理。医学随处显示着个体变异,这些变异组成了丰富多样的统计资料,也提出永无止境的研究课题,于是又促使多得眼花缭乱的统计方法的不断创新和建立,这是医学统计学本身的性质所决定的。

学习医学统计学,应该牢固树立起生物性个体变异的观念;各种医学指标独特的和分类的观念;抽样误差不可避免及各种条件下样本具有不同误差的观念;以及各种研究对象和研究方式含有不同变异的观念。

医学统计学是医学研究的重要工具,在相同的基础上对样本信息进行比较、分析、概括,依据概率通过逻辑推理作结论,属于从个别到一般的归纳推理型思维。这在一定程度上,与人们在其它学科学习和日常生活中养成的确定性的、偏于演绎推理型的思维方法有所不同。正因为此,统计学上得到的结论都具有概率性,它不能证明什么,但可以提供结论成立或不成立的概率,从而提高研究者的分辨能力,为科学决策提供依据。正确应用统计方法,能帮助我们正确认识客观事物,阐明事物的固有规律,从而把感性认识提高到理性认识。统计学不是万能的,它决不会改变事物的本来面貌,把原本不存在的结论“创造”出来。有些人在进行试验之前没有考虑充分,收集了一些不准确、不可靠或不全面的资料,而希望用统计方法来弥补,这是不可能的。医学统计学的全部功能仅在于帮助我们认识客观上存在的规律。

医学生学习医学统计学,并非要成为医学统计学的专业人才,其目的在于建立起统计学观念,学会从不确定性、机遇、风险和推断的角度去思考医学问题,尤其是医学研究问题;学会结合专业作出严密的试验设计并获得可靠、准确、完整的资料;学会运用统计方法充分挖掘资料中蕴含的信息,恰如其分地进行理性概括,并据此写出具有科学认证的研

究报告和学术论文。通过培养医学统计学思维，提高自身的科学素质和医学研究能力。

第二章 个体变异与变量分布

本章介绍个体变异的概念和统计资料的描述,包括指标的频数分布及其特征、常用描述指标,以及正态分布、二项分布、Poisson 分布等。

2.1 个体变异

变异(variation)是以具有同质性的观察单位为载体,某项观察指标在其观察单位之间显示的差别。由于观察单位通常即观察个体,故变异亦称个体变异(individual variation)。变异表现为定量的,如血清总胆固醇的测定值,可形成定量资料;变异表现为定性的,如病人性别,可形成定性资料;变异表现为等级的,如心功能分级,可形成等级资料。

变异是生物体在一种或多种不可控因素(已知的和未知的)作用下所产生反映的综合表现。就每个观察单位而言,其观察指标的变异是不可预测的,或者说是随机的。观察指标的表述用变量(variable),或称随机变量(random variable)。当观察值的个数(样本含量)达到足够多时,其分布将趋于稳定,并最终服从于总体分布。

变异现象广泛存在于人体及其他生物体。从这个意义上说,变异是医学研究中必须运用统计指标并进行统计分析的缘由。医学统计学的主旨就是运用变异的规律对医学资料进行统计分析。

2.2 频数分布

由于个体变异的普遍存在,医学研究中任何观察指标在各个个体上的观察结果是有不同的,当个体数增多时,这些观察结果将会呈现出一定的分布(distribution)规律。不同的观察指标有各自固有的分布规律,并依此作为选择相应的统计分析方法的重要依据。医学研究中观察到的原始数据(raw data)通常需要经过整理,用适当的形式表达其分布之特征。整理数据最有效的形式是频数分布(frequency distribution),根据频数分布(样本)可以初步判断指标分布(总体)的特征。用频数分布表示的数据称为分组资料(grouped data)。在样本含量足够大时,频数分布将接近其理论分布。

2.2.1 计量资料的频数分布

先看一个例子。

例 2.1 某市 1997 年 12 岁男童 120 人的身高(cm)资料如下。

142.3	156.6	142.7	145.7	138.2	141.6	142.5	130.5	134.5	148.8
134.4	148.8	137.9	151.3	140.8	149.8	145.2	141.8	146.8	135.1
150.3	133.1	142.7	143.9	151.1	144.0	145.4	146.2	143.3	156.3