



21世纪复旦大学研究生教学用书

复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列



Gonggong Weisheng yu Yufang Yixue Xil

公共卫生与预防医学系列

# 临床医学研究设计 和数据分析

● 主编 赵耐青

复旦大学出版社

## 公共卫生与预防医学系列已出书目：

医学人口学（第2版）

医用统计方法（第2版）

健康教育学（第3版）

临床试验

临床医学研究设计和数据分析



复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列

ISBN 7-309-04485-1



9 787309 044850 >

R·891 定价：45.00元



21世纪复旦大学研究生教学用书

129748

R4  
ZNR  
C1

复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列 复旦博学·公共卫生与预防医学系列



Gonggong Weisheng yu Yufang Yixue Xili

公共卫生与预防医学系列

# 临床医学研究设计 和数据分析

● 主编 赵耐青

主 审 曹素华  
 主 编 赵耐青  
 副主编 (按拼音顺序排列)  
 陈 峰 邓 伟 刘晓云 张文彤  
 周 琴

编 者 (按拼音顺序排列)

陈 峰	南京医科大学	邓 伟	复旦大学
董 伟	复旦大学	郜艳辉	广州中医药大学
郝元涛	广州中山大学	何艳青	复旦大学
刘晓云	复旦大学	罗剑锋	复旦大学
邱 宏	上海中医药大学	宋艳艳	上海第二医科大学
汪 涛	复旦大学	张文彤	复旦大学
赵耐青	复旦大学	赵新平	复旦大学
周 琴	复旦大学	朱 骥	复旦大学

解放军医学图书馆[书]



\*C0246677\*

复旦大学出版社

**图书在版编目(CIP)数据**

临床医学研究设计和数据分析/赵耐青主编. —上海:复旦大学出版社, 2005. 7

(博学·公共卫生与预防医学系列)

ISBN 7-309-04485-1

I. 临… II. 赵… III. ①临床医学-研究方法②临床医学-数据-分析 IV. R4

中国版本图书馆 CIP 数据核字(2005)第 030685 号

**临床医学研究设计和数据分析**

**赵耐青 主编**

---

**出版发行** 复旦大学出版社

上海市国权路 579 号 邮编 200433

86-21-65118853(发行部) 86-21-65109143(邮购)

fupnet@fudanpress.com <http://www.fudanpress.com>

---

**责任编辑** 王龙妹

**总编辑** 高若海

**出品人** 贺圣遂

---

**印刷** 江苏句容市排印厂

**开本** 787 × 1092 1/16

**印张** 23 插页 2

**字数** 532 千

**版次** 2005 年 7 月第一版第一次印刷

**印数** 1—3 000

---

**书号** ISBN 7-309-04485-1/R · 891

**定价** 45.00 元

---

如有印装质量问题, 请向复旦大学出版社发行部调换。

版权所有 侵权必究

## 内 容 提 要

本书为临床研究生和其他医科研究生的医学统计学教材。内容全面、丰富，不仅包括了医学统计的基本概念、计量和计数资料的描述性统计与推断性统计方法和研究设计，并且包括一些统计方法学的新进展。

为帮助读者较快地熟练掌握研究设计和数据统计分析的技能，本书以研究设计和数据统计分析为主要目标，介绍各种常用的研究设计，借助数据统计分析的案例介绍相关的统计分析方法。

本书借助运用统计软件Stata，淡化计算和相关的统计公式，强化实用性和基本概念。在编写过程中参考国外教材，书中还介绍了数据分析中的Stata操作。本书适合医科研究生，七年制、八年制临床专业学生和其他医学专业本科生用。

主 审 曹素华

主 编 赵耐青

副主编 (按拼音顺序排列)

陈 峰 邓 伟 刘晓云

张文彤 周 琴

编 者 (按拼音顺序排列)

陈 峰 南京医科大学

邓 伟 复旦大学

董 伟 复旦大学

郜艳辉 广州中医药大学

郝元涛 广州中山大学

何艳青 复旦大学

刘晓云 复旦大学

罗剑锋 复旦大学

邱 宏 上海中医药大学

宋艳艳 上海第二医科大学

汪 涛 复旦大学

张文彤 复旦大学

赵耐青 复旦大学

赵新平 复旦大学

周 琴 复旦大学

朱 骥 复旦大学

# 序

与疾病的抗争是人类发展史的重要组成部分。对疾病干预是医学的重要内容,对干预前后疾病的表现有确定性现象与随机现象。确定性现象是指在一定条件下必然发生的现象。正确认识疾病过程中的确定性现象与随机现象,是临床研究的目 的。患病的机体本身也在与疾病进行抗争,疾病的过程与表现是一个复杂的主观与客观的现象。在疾病发生与发展的整个过程中,不是每个患者都呈现“每况愈下”的局面。临床医师采取针对疾病相应的干预措施,为能评价临床干预措施,需对整个医疗过程疾病的现象进行观察与判断,此过程形成了临床医学研究。

《临床医学研究设计和数据分析》注重统计思想和统计概念,以实际应用为主要视角,按照研究设计分类介绍,借助统计软件 Stata,淡化统计计算过程,注重基本概念在实践应用中的作用,强化研究设计和掌握实际研究中的数据分析方法以及分析结果的解释,努力使学生能较快具备实际应用的能力。

本书力求从应用出发,帮助临床工作者揭示疾病发生与发展过程中的规律性,故也还可作为从事临床研究的科技工作者的参考书。

复旦大学公共卫生学院院长

姜庆五

2005年4月8日

# 前 言

研究设计和数据统计分析是医学科学研究中的一个重要组成部分。目前,许多国际性医学研究项目均需医学统计学人员参加。我国的《药品注册管理办法》规定新药临床试验必须自始至终有统计学人员参与。目前医学统计学是医学各专业本科生和研究生的必修课程。

本教材努力与国外教材接轨:突出统计基本概念和统计思想,注重实际应用的需要,以研究设计和数据统计分析为主线,介绍统计分析方法,结合计算机统计软件的应用,淡化统计计算,努力使本教材内容和编排更贴近医学科研的实际情况。本教材适用的对象是七年制、八年制临床专业学生和临床研究生。

Stata 软件是最适合于教学的一个统计软件,该软件功能强大,操作简便,输出结果针对性极强,并且许多功能都是针对医学研究背景设计的。根据我们的教学实践经验,只要安排 1 个学时介绍 Stata 的 4 个窗口和资料输入方法,就可以让学生根据书中的 Stata 简介,用 Stata 软件独立完成本书例题与习题的有关统计计算和作图,只需提供上机条件而无需另行安排上机课。有关 Stata 软件的更详细介绍,读者可以参阅陈峰教授编写的《现代医学统计方法与 Stata 应用》第二版。

为了使读者更有效地使用本教材,我们特地编制一些计算机辅助教学的模拟程序以及其他功能的统计程序,并撰写了《Stata 统计软件使用入门》。为了降低读者购书的成本,这些内容将不出现在本书内容中,读者可以登陆《医学统计之星》网站或 <http://www.shmu.edu.cn>(课程下载)下载和更新这些内容。

本教材的编写得到复旦大学各级领导,尤其是复旦大学研究生院的大力支持。本教材出版得到复旦大学出版社的领导和老师的大力支持和帮助。



目  
录

第一章 绪论	1
第一节 统计学中的几个基本概念	1
第二节 医学研究中的统计问题和设计问题	5
第三节 学习数据统计分析和研究设计应注意的问题	6
第四节 小结	7
习题	7
第二章 统计描述	9
第一节 计量资料的统计描述	9
第二节 分类资料的统计描述	21
第三节 常用统计图表	27
第四节 小结	36
习题	36
第三章 概率分布	39
第一节 二项分布	39
第二节 Poisson 分布	42
第三节 正态分布	44
第四节 小结	51
习题	52
第四章 总体均数的估计和假设检验	54
第一节 抽样误差	54
第二节 $t$ 分布	56

第三节	样本率的分布 .....	57
第四节	参数估计 .....	58
第五节	假设检验 .....	61
第六节	二类错误 .....	65
第七节	小结 .....	66
	习题 .....	67
<b>第五章</b>	<b>研究概述 .....</b>	<b>71</b>
第一节	实验设计的基本概念 .....	71
第二节	常见的医学研究设计简介 .....	73
第三节	研究设计常见的问题和注意事项 .....	81
第四节	小结 .....	82
	习题 .....	82
<b>第六章</b>	<b>完全随机设计两样本资料的统计分析 .....</b>	<b>84</b>
第一节	连续变量的两样本资料检验 .....	84
第二节	离散型变量的两样本资料检验 .....	93
第三节	小结 .....	103
	习题 .....	104
<b>第七章</b>	<b>完全随机设计多样本资料的统计分析 .....</b>	<b>107</b>
第一节	完全随机设计资料的方差分析 .....	108
第二节	多组比较的秩和检验 .....	113
第三节	行列表资料统计分析 .....	117
第四节	小结 .....	120
	习题 .....	120
<b>第八章</b>	<b>随机区组设计和资料分析 .....</b>	<b>123</b>
第一节	随机区组设计 .....	123
第二节	配对设计资料的统计分析 .....	124
第三节	随机区组设计资料的统计分析 .....	130
第四节	小结 .....	136
	习题 .....	137

<b>第九章 析因设计和资料分析</b> .....	139
第一节 析因设计.....	139
第二节 析因设计资料的方差分析.....	141
第三节 小结.....	149
习题.....	150
<b>第十章 相关与回归</b> .....	152
第一节 相关分析.....	152
第二节 直线回归分析.....	157
第三节 多重线性回归.....	165
第四节 Logistic 回归 .....	167
第五节 小结.....	173
习题.....	173
<b>第十一章 生存分析</b> .....	177
第一节 基本概念.....	177
第二节 生存曲线估计.....	179
第三节 生存曲线的比较.....	185
第四节 Cox 回归 .....	189
第五节 多结局生存分析回归模型.....	193
第六节 小结.....	198
习题.....	198
<b>第十二章 随机对照试验和数据分析</b> .....	201
第一节 随机对照试验.....	201
第二节 重复测量资料的统计分析方法.....	205
第三节 协方差分析.....	218
第四节 小结.....	222
习题.....	222
<b>第十三章 抽样调查研究</b> .....	224
第一节 调查研究概述.....	224
第二节 横断面研究的设计和分析.....	227

第三节	病例-对照研究的设计和分析 .....	233
第四节	小结 .....	239
习题	.....	240
<b>第十四章</b>	<b>临床诊断试验 .....</b>	<b>241</b>
第一节	试验设计中的基本概念 .....	241
第二节	常用诊断试验的评价指标 .....	242
第三节	ROC 曲线的应用 .....	245
第四节	小结 .....	252
习题	.....	253
<b>第十五章</b>	<b>临床试验 .....</b>	<b>255</b>
第一节	临床试验概况 .....	255
第二节	临床试验方案设计及其特殊问题 .....	256
第三节	病例报告表和数据管理 .....	260
第四节	小结 .....	263
习题	.....	264
<b>第十六章</b>	<b>交叉设计 .....</b>	<b>265</b>
第一节	交叉设计的特点 .....	265
第二节	交叉试验资料的分析 .....	266
第三节	应用中的注意事项 .....	272
第四节	小结 .....	272
习题	.....	272
<b>第十七章</b>	<b>循证医学与 Meta 分析简介 .....</b>	<b>273</b>
第一节	循证医学 .....	273
第二节	Meta 分析简介 .....	275
第三节	小结 .....	291
习题	.....	291
<b>第十八章</b>	<b>量表的信度和效度 .....</b>	<b>292</b>
第一节	效度、信度的概念 .....	292

第二节 实例分析.....	295
第三节 小结.....	299
习题.....	300
<b>第十九章 Bayes 后验统计推断简介</b> .....	<b>301</b>
第一节 Bayes 统计学的基本概念.....	302
第二节 Bayes 后验统计推断实例.....	304
第三节 小结.....	310
习题.....	310
参考文献.....	311
附录 1 统计用表.....	313
附表 1 标准正态分布曲线下的面积, $\Phi(-u)$ 值.....	313
附表 2 $t$ 界值表.....	314
附表 3 $F$ 界值表(方差齐性检验用).....	316
附表 4 $F$ 界值表(方差分析用).....	319
附表 5 $q$ 界值表(SNK 法用).....	324
附表 6 Dunnet $t$ 检验 $q'$ 界值表.....	325
附表 7 百分率的可信区间.....	327
附表 8 $\chi^2$ 分布界值表.....	333
附表 9 Poisson 分布 $\mu$ 的可信区间.....	334
附表 10 $T$ 界值表(配对比较的符号秩和检验用).....	335
附表 11 $T$ 界值表(两样本比较的秩和检验用).....	336
附表 12 $H$ 界值表(三样本比较的秩和检验用).....	338
附表 13 随机单位组设计秩和检验的界值表.....	338
附表 14 相关系数 $r$ 界值表.....	339
附表 15 Spearman 相关系数 $r_s$ 界值表.....	340
附录 2 数据分析的统计方法选择小结.....	342
附录 3 汉英词汇对照.....	346

医学科学研究中的大多数观察结果都是不能事先确定的,即使条件完全相同的两次观察结果往往也是不同的,观察结果具有随机性。在概率论中称结果具有随机性的观察为随机试验,每种可能的结果均称为随机事件,经常用英文大写字母 A、B、C 表示。如观察某一医院某一医生用某药抢救卒中(中风)患者这一现象,被救活是一个可能的结果,因抢救失败而死亡则是另一个可能的结果,抢救前不能确定何种结果将会出现,故被救活是一个随机事件,并可以用 A 表示。如果抢救的结果为患者被救活了,则称随机事件 A 发生了,抢救失败而死亡则称 A 未发生。在医学研究领域中的研究设计、资料收集和结果分析常常需要运用统计学知识,医学统计学(medical statistics)就是统计学原理和方法在医学研究领域中的应用。20 世纪中期以后,医学统计学逐渐形成一门学科,其在医学研究中的作用也愈显重要。目前,许多国际性医学研究项目均需医学统计学人员参加。我国的《药品注册管理办法》规定新药临床试验必须自始至终有统计学人员参与。目前医学统计学已经成为医学研究领域中的重要组成部分,并是医学各专业本科生和研究生的必修课程。

### 第一节 统计学中的几个基本概念

为了帮助读者较轻松地了解统计学的一些基本概念,我们将通过介绍医学研究实际例子,引出相关的概念,然后归纳其中的一些常用统计学基本概念。

**[例 1.1]** 研究 1998 年某地 7 岁男孩的身高情况。假如该地在 1998 年的 7 岁男孩有 10 万人,则最直接的方法就是普查:测量这 10 万个儿童的身高,然后进行统计分析,但这样做的工作量非常大。也可以用抽样调查的方法,从 10 万个儿童中随机抽取 200 个儿童(这里的“随机”是指 10 万个儿童均有相同的机会被抽取),测量他们的身高,通过分析这 200 个儿童的身高,推断该地 10 万个 7 岁男孩身高情况。在这个研究中所涉及的基本概念如下。

研究目的:研究 1998 年某地 7 岁男孩的身高情况。

研究对象:1998 年在该地的 10 万个 7 岁男孩。

观察单位(又称为个体):研究对象中的每个 7 岁男孩。

观察指标(又称为变量):身高。身高的测量值(观察值)又称为身高变量取值或身

高变量值。

总体:该地 1998 年的 10 万个 7 岁男孩身高观察值的全体(即 10 万个身高观察值构成的一个集合)。

样本:随机抽取的 200 个 7 岁男孩身高观察值构成的集合称为样本。

该研究通过分析 200 个 7 岁男孩的身高,从而了解 10 万个 7 岁男孩身高情况,这样的研究称为抽样研究。即抽样研究的目的是通过样本信息,了解总体情况。

这 10 万个 7 岁男孩的身高平均数在这个研究中称为该地区 7 岁男孩身高的总体均数,这 200 个 7 岁男孩的身高平均数在这个抽样研究中称为身高的样本均数。可以用抽样研究所得到的样本均数估计总体均数。

对于一般的研究,引入下列基本概念。

1. 个体 个体(individual)是统计分析根据研究目的所确定最基本的研究对象单位,所以个体又称为**观察单位**。根据不同的研究目的,个体可以是一个人、一只大鼠、一个家庭、一个地区、一个检测样品、一个采样点等。例如,观察单位是一个人,则 100 个观察单位就是 100 个人;又如观察单位是检测样品,则 50 个观察单位就是 50 个检测样品。

2. 同质 具有相同性质的观察单位称为**同质**(homogeneous),否则,称为**异质**(heterogeneous)。例如调查某地 1995 年正常成年女子的糖化血红蛋白(HbA1C),则研究对象是该地 1995 年的正常成年女子,观察单位是每个女子,观察指标是糖化血红蛋白,观察值是每个人的糖化血红蛋白测量值,同一地区、同一年份、同为正常成年人和同为女性构成了研究对象同质的要素。同质的要求与研究目的有关,如调查某地 1995 年正常成年女子的雌激素水平,尽管研究对象是同一地区、同一年份、同为正常成年人和同为女性,由于女性在绝经后的雌激素水平有较大下降,如果研究者把绝经和未绝经的研究对象混合在一起,并且不加区分,则对于研究雌激素水平而言,这些研究对象显然是异质的。

3. 总体 总体(population)是指同质的所有个体某指标观察值(测量值)的集合。由于在实际研究中,往往需要观察或测量多个指标,而这些指标之间往往伴有某种关联,故多个观察指标构成了个体的一组观察指标。为了叙述方便,往往简单地称总体是根据研究目的确定同质个体的全体。总体分为有限总体(finite population)和无限总体(infinite population)。有限总体中个体总数是有限的;无限总体中个体总数是无限的。在医学研究中,许多研究目的所对应的总体在严格意义上是有限总体,但是这些总体中的个体总数往往非常大,与无限总体在实际应用中几乎无差别,所以把个体总数非常大的有限总体近似视为无限总体。

4. 样本 在一个较大范围的研究对象中随机抽出一部分个体对某些研究指标进行观察或测量,这些个体研究指标的测量值构成的集合称为**样本**(sample),样本中的个体总数称为**样本量**(sample size)。由于在实际研究中,往往需要观察或测量多个研究指标,为了叙述方便,往往简单地称样本是由一部分个体构成的。

5. 变量 亦称观察指标。变量(variable)的取值表示观察值(或测量值)或对应的观察结果,亦称**资料**(data)。如:身高测量值;尿常规检查结果可以为-、+、++;血型为 A、B、AB、O。由于医学观察结果往往在观察前是未知的,而且即使在相同条件下的重复观察结果往往也是不同的(如观察血压等),故称这种观察结果是随机的,并且称相应的

观察指标为随机变量,但在医学统计中常简称为变量。变量可以分为连续型变量(continuous variable)和离散型变量(discrete variable)。

1) 连续型变量可以在一个区间中连续取值,即在忽略测量精度的情况下,连续型变量在理论上可以取到区间中的任意一个值,并且通常含有测量单位。观察连续型变量所得到的数据资料称为计量资料(measurement data),也可以称为定量资料。如例 1.1 中的身高变量就是连续型变量,身高资料为计量资料。

2) 离散型变量取值范围是有限个值或者一个数列构成的。表示分类情况的离散型变量又称为分类变量。观察分类变量所得到的资料称为分类资料。分类资料可以分为二分类资料和多分类资料,而多分类资料又分成无序分类资料和有序分类资料。

(1) 二分类资料:如症状指标分为感染或未感染,疗效指标为有效或无效;又如性别变量为男或女等。观察可能的结果只有两个,记录这种观察结果的资料称为二分类资料。通常用变量取值为 0 和 1 对应两种可能的结果,所以这种变量称为 0—1 变量,其相应的资料称为 0—1 资料。

(2) 多分类有序资料:如病情指标分为无症状、轻度、中度和重度;又如疗效可以分为无效、进步、有效和痊愈。观察可能的结果只有若干个,并且这种观察结果在背景意义上含有程度或等级的意义,通常用变量取值为 1, 2,  $\dots$ ,  $m$  对应表示相应  $m$  个程度或等级,所以这种分类变量称为有序的,而这种分类变量取值亦称有序分类资料。有序分类资料可以按类汇总,统计每个等级的个体数,并将按类汇总的统计结果编制成表格形式的资料,这种汇总后的资料又可称为等级资料。

(3) 多分类无序资料:如血型可以分为 A、B、AB 和 O 型;又如慢性气管炎可以分为单纯型、喘息型、单纯型合并肺气肿和喘息型合并肺气肿。观察可能的结果只有若干个并且这种观察结果在背景意义上没有程度或等级的含义,通常用变量取值为 1, 2,  $\dots$ ,  $m$  对应表示相应  $m$  个无序分类,故这种变量的取值没有大小的背景意义,仅是指示分类的作用,所以这种分类变量称为无序的,其取值亦称无序分类资料。无序分类资料可以先按类汇总,分别统计每一类的个体数,并将按类汇总的统计结果编制成表格形式的资料,这种汇总后的资料又可称为计数资料。

(4) 特殊计量资料:有些观察指标,例如白细胞计数,其取值虽然是离散的,但不具有分类的性质,因此通常把这类观察指标的资料作为较为特殊的计量资料。

6. 频率 考察某个可能的观察结果(即随机事件),记为 A。若在  $n$  次观察中,随机事件 A 发生了  $m$  次,则称随机事件 A 发生的比例  $f = \frac{m}{n}$  为频率(frequency),  $m$  称为频数。显然有  $0 \leq f \leq 1$ 。例如在治疗幽门螺杆菌感染转阴情况的研究中,把治疗一个幽门螺杆菌阳性的患者视为一次观察,患者通过治疗后幽门螺杆菌转阴了则定义为事件 A 发生了一次,反之,定义为事件 A 未发生。治疗了  $n$  个幽门螺杆菌感染的患者,其中有  $m$  个人治愈(幽门螺杆菌呈阴性),也就是观察了  $n$  次,事件 A 发生了  $m$  次,因此该事件 A(幽门螺杆菌治愈)的发生频率  $f = \frac{m}{n}$ , 并称为  $f = \frac{m}{n}$  为幽门螺杆菌的治愈率。如在医学研究中,如患病率、病死率等都是频率。频率  $f$  是一个随机抽样观察的结果,因此频率  $f$  呈



一定的随机波动。如在某地区随机抽样调查糖尿病的患病率,其结果如下(表 1-1)。

表 1-1 某地区随机抽样调查糖尿病的患病率

抽样(调查)人数( $n$ )	100	500	1 000	5 000	10 000	50 000	100 000	1 000 000
糖尿病人数( $m$ )	12	48	102	493	992	4 999	10 003	99 999
频率(患病率%)( $f$ )	12.00	9.60	10.20	9.86	9.92	10.00	10.00	10.00

由表 1-1 可以看到频率  $f$  呈某种随机性。但随着抽样人数  $n$  的增大,频率(患病率) $f$  随机波动的幅度越来越小并且趋向常数 10%。可以证明:当观察次数  $n$  越来越大,频率  $f$  的随机波动幅度越来越小,并最终趋向于一个常数,这个常数被定义为随机事件  $A$  发生的概率(又称为概率的统计定义)。

7. 概率 概率(probability)刻画随机事件  $A$  发生可能性大小,其取值界于 0 和 1 之间,记为  $Pr(A)$ 。随机事件  $A$  发生的可能性越小,概率  $Pr(A)$  就越接近 0;随机事件  $A$  发生的可能性越大,概率  $Pr(A)$  越接近 1。不可能事件发生的概率等于 0,必然事件发生的概率等于 1。如在医学研究中,往往用人群的患病率估计每个人患病的概率;用患病人群的治愈率估计每个患者被治愈的概率等。

在统计学中,如果随机事件发生的概率  $\leq 0.05$ ,则可以认为该事件是一个小概率事件,表示该事件在大多数情况下不会发生。一般认为小概率事件在一次随机抽样中不会发生,这就是小概率原理。小概率原理是统计推断的基础。

8. 总体参数 刻画总体特征的指标称为总体参数(parameter)。例如某研究的总体为全体正常人,则所有正常人中抗乙型肝炎病毒表面抗原抗体(HBsAb)呈阳性的比例称为 HBsAb 的总体阳性率。再如总体中某个指标所有个体变量值的平均数称为总体均数。

9. 统计量 刻画样本特征的统计指标称为统计量(statistic),例如样本中某个指标的所有观察值的平均数称为样本均数;如某样本中 HBsAb 呈阳性的比例称为 HBsAb 的样本阳性率。在实际工作中,总体参数往往是未知的,这些未知总体参数可以用样本统计量进行估计。例如用样本均数估计总体均数、用 HBsAb 的样本阳性率估计 HBsAb 的总体阳性率等。

10. 个体变异和资料分布 例如同为正常发育的 7 岁男孩,各人体重各不相同;又如,病情相同的患者服用相同的药物,其疗效也不尽相同,这种个体之间的差异称为个体变异(individual variation)。更严格地说,同质个体的某指标之间的差异称为个体变异。在生物医学研究领域里,个体变异是普遍存在的,即使在完全相同的自然条件下生长的两个生物体也存在差异。由于在观察或测量前都不能事先确定这些个体变异的大小,因此这些个体变异是随机的。个体变异在概率意义下是有规律的,这种随机变异的规律性表现为观察值出现在不同范围中的概率,所以称这种随机变异的规律性为该观察指标(变量)取值的概率分布,简称为资料分布(data distribution)。例如,在某地区抽了 10 000 个 10 岁健康男孩,并测量其身高,结果为身高  $< 125$  cm 共有 720 人(占 7.2%);身高在 125~135 cm 共有 8 950 人(占 89.5%);身高  $> 135$  cm 共有 330 人(占 3.3%),可以发现该地 10 岁健康男孩的身高规律:身高  $< 125$  cm 的概率约为 7.2%、身高 125~135 cm 的