本书由复旦大学出版基金资助出版

流行病学方法与模型

姜庆五 陈启明 编著

复旦大学出版社
www.fudanpress.com

本书由复旦大学出版基金资助出版

流行病学方法与模型

Epidemiological Methods and Model

姜庆五　陈启明　编著

袁鸿昌　审阅

復旦大学出版社

图书在版编目(CIP)数据

流行病学方法与模型/姜庆五,陈启明编著. —上海:复旦大学出版社,2007.9
ISBN 978-7-309-05535-1

Ⅰ. 流… Ⅱ. ①姜…②陈… Ⅲ. 流行病学 Ⅳ. R18

中国版本图书馆 CIP 数据核字(2007)第 071574 号

流行病学方法与模型

姜庆五 陈启明 编著

出版发行 复旦大学出版社 上海市国权路 579 号 邮编 200433
86-21-65642857(门市零售)
86-21-65100562(团体订购) 86-21-65109143(外埠邮购)
fupnet@fudanpress.com http://www.fudanpress.com

责任编辑 傅淑娟
总 编 辑 高若海
出 品 人 贺圣遂

印 刷 上海第二教育学院印刷厂
开 本 787×1092 1/16
印 张 31.375
字 数 767 千
版 次 2007 年 9 月第一版第一次印刷
印 数 1—2 100

书 号 ISBN 978-7-309-05535-1/R·982
定 价 68.00 元

前　言

流行病学研究疾病在人群中的现象，用数学语言与数学模型概括疾病在人群的表现，被称为理论流行病学或者数学流行病学。理论流行病学是流行病学工作者对疾病研究的一个境界，只有当疾病在人群的现象被人们认识得非常清晰的时候，流行病学家才会考虑应用数学模型去概括疾病在人群中的现象。当代流行病学能够应用数学模型去表述疾病在人群中的表现，并应用数学模型进行推理、演绎、证明疾病在人群中的自然现象与社会现象，是流行病学家对疾病认识的一个深刻阶段。

数学模型不同于单纯的观测，它具有逐级抽象的特点。客观实际、现实世界中的抽象只是数学模型的初级抽象。流行病学工作者在流行病学的调查过程中便完成初级抽象，用各种相应的指标来描述疾病在人群中的现象。此阶段我们称为分析流行病学。用数学模型将疾病在人群中的复杂现象用数学语言进行概括，是离开具体事和物的疾病的数量关系和空间形式的流行病学模型，是数学的高级抽象。流行病学模型的研究对象是一种形式化的思想材料，是流行病学工作者经过加工了的思想，是对疾病现象的高度概括与深刻的认识。数学模型的验证、推广是将发展的结果以演绎推理的形式系统化、逻辑化。流行病学的数学模型帮助我们理解与回答疾病在人群中传播的复杂流行病学问题，此阶段我们称为理论流行病学。

复旦大学公共卫生学院（原上海医科大学公共卫生学院）重视应用数学模型的研究。我国著名的流行病学家苏德隆教授开拓性地应用数学模型为流行病学的研究树立了典范。苏德隆教授认为数学与统计学都是流行病学的重要工具。1950 年他修正了 Reed-Frost 模型，成功地将人群的免疫现象引进了疾病的传播模型。他建立的钉螺的负二项分布模型，提出钉螺分布的不均匀性，并将此理论推广至疾病的控制，从模型上验证了对疾病的控制是必须反复斗争的，否则将会“死灰复燃，前功尽弃”。苏德隆教授的成就推动了他的同事与学生继续在理论流行病学方面的研究，徐志一教授的肝炎控制模型研究、沈福民教授在遗传流行

前言

病学模型的研究、袁鸿昌教授在血吸虫病传播与控制模型的研究、俞顺章教授在肝癌病因模型的研究、龚幼龙教授在结核病控制模型研究等方面的成就也都为今天复旦大学公共卫生学院在流行病学疾病模型研究方面作出了贡献与奠定了基础。本书作者是在这样一个生态环境下开展的研究，深感本书的形成是参天大树上的一颗果实。

本书的启动，是作者参与了袁鸿昌教授主持的科技部“七五”、“八五”攻关课题“血吸虫病流行因素与传播规律的研究”等项目，在“十五”期间，参与了李立明教授主持的国家自然科学基金重点项目“流行病学与生物统计学结合进行因果探索的理论方法研究”，以及目前正在参与的曹务春教授主持的自然科学基金重大课题“基于现代信息技术研究传染病时空传播与流行规律”。这些研究项目促使作者有时间系统思考近年来复旦大学公共卫生学院流行病学教研室在疾病传播模型上的研究成果。当然，作者还想以本书作为课题组目前正在承担的自然科学基金重大课题的子课题“不同时空尺度的血吸虫病流行规律研究及其模拟与分析”的一部分研究结果。本书的内容中还应用了复旦大学公共卫生学院流行病学博士研究生论文的部分研究结果，如郜艳晖副教授的“疾病家庭相关测量模型研究”、蔡全才副教授的“SARS 传播模型研究”、余金明教授的“血吸病感染模型的研究”以及于淑丽博士的“乙型肝炎传播与控制模型研究”。在本书编写过程中，得到了袁鸿昌教授的鼓励，本书的体系形成无不包含了袁鸿昌教授的精心指导，特别感谢袁鸿昌教授对全书的评阅。

在本书出版时，正值上海医科大学 80 周年诞辰，作者的成长是与上海医科大学的发展分不开的。因此，愿将本书作为向上海医科大学 80 周年华诞奉献的一份礼物。

姜庆五　陈启明

2007 年 3 月

目录

第一篇　流行病学研究设计

第二篇　流行病学计数资料分析

第三篇　流行病学计量资料分析

第四篇　病因研究与推断

第五篇　疾病流行模型

第一篇

流行病学研究设计

- 第一章　流行病学资料与分析方法
- 第二章　流行病学调查设计

第一章

流行病学资料与分析方法

流行病学研究疾病在人群中的现象。疾病在人群中的发生不是随机的，每个人发生疾病的概率是不等的，疾病是否发生取决于个体的功能与其环境。一个人群中发生疾病的频率可以通过人群中的个体的观察进行估计。流行病学研究的目的是通过对人群的干预而达到对疾病的预防与控制。对人群疾病与疾病相关因素的调查，形成了流行病学研究的基础。

第一节 流行病学资料及其分类

流行病学资料是指流行病学现场调查收集或实验室试验观测得到的对象标志状态值数据或试验结果数据，可从流行病学研究的目的与方法、资料的属性、资料对应的变量维数以及资料对应的变量或向量间关系等不同角度进行分类，以便采用对应有效的分析方法，获取应得的合理结论。

一、按流行病学研究目的与方法分类

流行病学研究疾病流行的群体现象和因果规律，借以预测并制订和评价相应的防治策略措施，有效控制疾病的发生和流行，减少疾病负担，提高生命质量。

流行病学研究的基本方法是流行病学调查（包括观测调查与试验调查，无干预调查与有干预调查）。不同研究目的有不同调查类别，如横断面（cross-section）调查、病例对照（case-control）调查、回顾性（retrospective）或前瞻性（prospective）随访或跟踪（follow-up）调查，以及它们的组合调查。归纳起来，与横向研究与纵向研究对应，可分为横向或横断面调查与纵向（longitudinal）调查两类。病例对照调查属有对照的“由果及因”横向调查。队列（cohort）与个案（single case）随访或跟踪调查属“由因及果”纵向调查。如图 1-1。

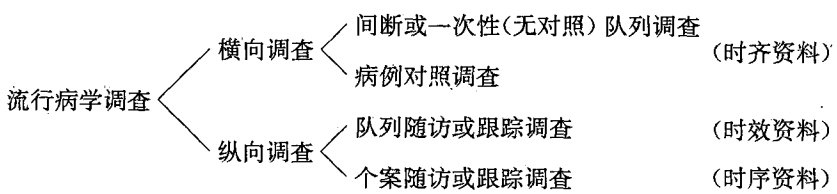

图 1-1 流行病学调查分类示意

按流行病学研究目的与方法，流行病学资料与流行病学调查对应，可分为横向或时齐（time point/period）资料、纵向（longitudinal/panel）或时效（time effective）资料以及时序（time order/series）资料：①由间断或一次性抽样调查（包括病例对照或回顾性与前瞻性调

查)得到的数据资料称为横向或时齐资料;②由队列随访或跟踪调查得到的数据资料称为纵向或时效资料;③由个案随访或跟踪调查得到的数据资料称为时序资料。

按时间顺序,横断面调查是无重复观测,随访或跟踪调查是有重复观测。所以,横向或时齐资料、纵向或时效资料及时序资料三者间的本征区别在:①横向或时齐资料是样本无重复观测资料;②纵向或时效资料是样本有重复观测资料(重复观测次数较少);③时序资料是个案观测(如区域人群定期性普查、临床病例持续性监测)值序列资料。序列较短的时序资料称动态数列(moving series),序列可无限延续的时序资料称时间序列(time series)。如图 1-2,(a)为时齐资料、(b)为时序资料、(c)为时效资料。

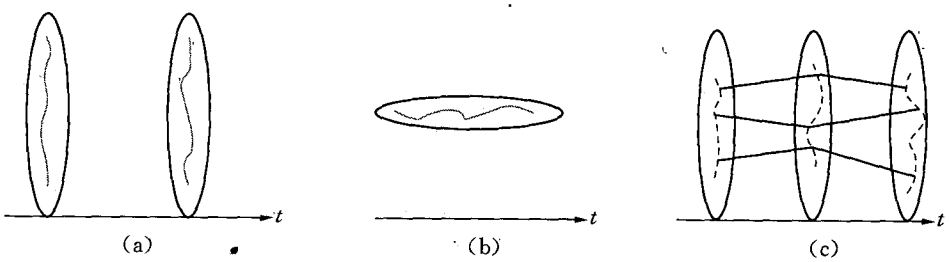

图 1-2 不同调查方法资料特点

流行病学调查的时间尺度随起始点选择不同有多种。表 1-1 所列是其经常用到的几种,它们间又常存在一定联系。例如,在许多研究中,时间的三种尺度(出生、年龄和日期)中任两种尺度均可确定第三种尺度。如由年龄与出生队列(birth cohort)便可确定日期。

表 1-1 几种常用时间尺度

起始点(starting point)	时间尺度(time scale)
出生(birth)	年龄(age)
任一固定日期(any fixed date)	日历日期(calendar time)
首先暴露(first exposure)	暴露时间(time exposed)
进入研究(entry into study)	研究时间(time in study)
发病(disease onset)	发病时间(time since onset)
开始治疗(start of treatment)	治疗时间(time on treatment)

二、按资料属性分类

无论时齐资料、时序资料还是时效资料,按其属性均可分为名义(nominal)变量(观测)值资料或计数资料与数值变量值资料。名义变量值资料或计数资料根据其所对应品质标志状态类的有序与无序又分无顺序分类资料(简称无序资料)与有顺序分类资料(简称有序资料)。数值变量值资料分离散型资料、连续型资料与非离散非连续型资料,如图 1-3。离散型资料是取值可数的数值变量调查观测或试验观测结果值资料,连续型资料是在取值区间内可取任意值的数值变量调查观测或试验观测结果值资料,非离散非连续型资料是取值既不可数又不连续的数值变量调查观测或试验观测结果值资料。

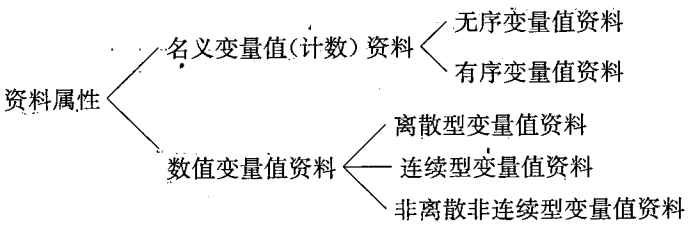

图 1-3 流行病学资料按属性分类

三、按资料对应的变量维数分类

按资料对应的变量维数分类，可分为单变量(标量)值资料与多变量(向量或矢量)值资料。标量值资料是与个体对应的单标志调查观测或试验结果值资料。向量或矢量值资料是与个体对应的多标志调查观测或试验结果值资料。例如，测量人的身高得到的是标量值资料。若同时测量人的身高、体重、脉搏和血压，则得到的是(四维)向量值资料。

四、按资料对应的变量或向量间关系分类

按资料对应的变量或向量间关系分类，可分独立变量值资料与相关变量值资料。独立变量是与其他变量没有任何关系的变量，独立变量值资料的取得不受其他变量值资料影响。相关变量是相互间有关联的变量。相关变量按相互间关系可区分为线性相关与非线性相关，按相互间作用可区分为说明(explanation)变量或预报(predictor)变量或回归(regressive)变量、协变量(covariant)与应(response)变量。相关变量值资料的取得相互间都有影响。

考虑到时间尺度对资料搜集的重要性与处理方法上的不同，流行病学资料的整理与描述和分析可按资料的时齐性、时序性与时效性区别进行。

第二节 流行病学资料的分析方法

疾病流行现象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，如血吸虫病粪检阳性必然有血吸虫感染(现象)。随机现象是指在一定条件下可能发生也可能不发生的现象，如感染了结核菌可能生结核病也可能不生结核病(现象)。

流行病学资料分析是旨在描述和揭示群体中疾病流行现象的数量变化规律。分析方法包括流行病学统计数学方法与流行病学动力学数学方法，统称流行病学数学方法。流行病学统计数学方法描述和揭示疾病流行现象的统计(随机)规律性，简称流行病学统计分析或流行病学统计方法(包括统计描述分析方法与统计推断分析方法)。流行病学动力学数学方法描述和揭示疾病流行现象的形成与变化过程(动力学规律性)，简称流行病学动力学方法或动力学模型。疾病流行现象的必然性(因果)规律寓于偶然性(随机)规律之中，在不断调查与实验的基础上，通过对流行病学(数据)资料的统计分析，逐步实现对疾病流行因果规律的定量(模型)化机理性认识。流行病学统计方法用于对流行病学资料的统计分析。流行病学动力学模型用于定量描述疾病流行的因果机理规律。

流行病学数学方法是现代流行病学研究疾病流行原理用的定量方法。

一、流行病学统计方法研究内容与研究方法

流行病学统计方法是生物统计(方法)的一门分支学科,是根据统计学理论与方法,以疾病流行的群体现象为研究对象,科学地设计流行病学调查或实验方案,定量地描述和揭示疾病流行的统计规律,预测和评价干预措施对疾病流行的控制功能和成效(cost and effective)。内容包括:流行病学调查或实验资料搜集(包括流行病学调查或实验设计);流行病学调查或实验资料整理与统计描述;以及流行病学资料的统计推断。统计描述与统计推断统称统计分析。

流行病学统计方法的研究方法是流行病学研究方法与统计研究方法的结合,即按流行病学原理,根据统计学理论组织与设计流行病学调查或实验,搜集和整理流行病学调查或实验资料,对疾病流行的群体现象表现规律进行统计描述和推断,对疾病干预成效进行统计评估与预测。其特点亦由流行病学研究方法的特点与统计研究方法的特点所定,既有观测所得数据资料的大量与随机的共同统计特性又有流行病学学科专业应用特性。例如,由流行病学调查或实验所得的无干预或有干预观测资料,往往有因素混杂,资料的偏性不限于统计的方法和技术,且所得的资料多为有关生物群体的资料,常与时间量有关,需要有别于通常使用的相应统计方法,反映了应用统计不单纯是统计方法的应用,同时也是统计方法内容的丰富与发展的一种动力。

二、流行病学统计方法基础概念

(一) 统计总体与个体

统计总体简称总体或群体(population),是与研究目的相应的研究对象的全体组成的集合。组成总体的每个对象称为个体。所以,总体与个体是同时可根据研究目的确定的。总体有有限总体与无限总体。由有限个个体组成的总体称有限总体,由无限个个体组成的总体称无限总体。组成总体的个体数目称为总体大小(size)或总体容量。如若研究某种疾病在一个有 100 万人口(30 万个家庭)的地区的流行(现象),研究对象是人,总体就是由该地区所有人口全体组成的人群,该地区内的每个人就是组成该群体中的个体,该群体是有限总体,大小或容量为 100 万。

(二) 抽样与样本

流行病学统计方法主要研究特点是了解部分认识总体。从总体中抽出一部分有代表性的个体进行观测的方法称为抽样调查方法或试验方法,简称抽样方法。被抽到的这一部分个体(组成的集合)称为样本(sample),组成样本的单位称为抽样单位(unit)或样品。组成样本的单位数称为样本大小或样本容量。组成样本的抽样单位与组成总体的个体单位有时一致有时不一致。例如研究疾病的流行,在上述有 100 万人口的地区抽取 1 000 人进行调查,抽样单位与组成总体的个体单位一致是人,样本大小为 1 000;若研究疾病流行的家庭聚积性,从由这 100 万人口组成的 30 万家庭中抽取 100 户家庭,这时抽样单位与组成总体的个体单位不一致,抽样单位为家庭,样本大小为 100。

抽样有统计抽样与非统计抽样。为使被抽到的个体具有代表性,通常采用统计抽样。统计抽样有简单随机抽样、分层随机抽样和多阶段随机抽样等不同随机抽样方法,基本的统计抽样方法是简单随机抽样,简称随机抽样(random sampling)。所谓简单随机抽样是能使总体中的每个个体(或抽样单位)被抽到的可能性大小都一样的抽样方法,这样抽到的具有统计代表

性的样本称为随机样本。例如，对总体中每个个体进行编号后用有重复（有放回）抽签法抽取部分个体组成样本的方法是一种简单随机抽样法，所得样本是对总体具有代表性的随机样本。

流行病学统计方法中的流行病学调查即指（流行病学）统计抽样（调查）。基本调查方法是横断面调查、病例对照检查以及（回顾性与前瞻性）队列调查。横断面调查是特定时期的统计调查，常用于疾病流行或暴发现况调查或发病原因的探索性人群调查。病例对照调查是回顾性人群分组（分病例组与对照组）调查，通过病例组与对照组暴露因素不同探索发病原因，人称由果究因调查。队列调查是不同暴露因素水平特定健康人群的回顾性与前瞻性追踪或随访调查，根据因素水平不同发病率不同探索发病原因，人称由因及果调查。根据不同调查目的与要求，可由这 3 种调查方法组合成其他调查方法。

（三）个体标志与总体指标

个体标志简称标志（symbol），是代表个体属性和特征的名称。标志的状态称为标志值。如人群中人的民族、性别、年龄、健康状况等是个体（人的）标志，民族的回、满、藏、苗、维吾尔、蒙古……汉族等，性别的男、女，年龄的 0、1、2……岁或如 0～15、16～45、46～60、61～岁年龄组，健康状况的良、中、差等分别是相应标志的状态即标志值。

标志有数量（quantity）标志和非数量标志。非数量标志亦称品质（quality）标志。品质标志有两状态（如性别）与多状态（如民族）分类或无序标志和分等级或有序标志（如健康状况）。数量标志的状态值有可数（如心率）和不可数（如血压）、有限（如体温）和无限（如细胞数、寿命）、间断（如脉搏、心率）和连续（如血压、体温、寿命）。品质标志和数量标志有区别又有联系。品质标志可数量化，数量标志可品质化。如性别用数 0、1 表示，体重用超重量级、重量级、次重量级、中量级、轻量级、次轻量级、羽量级等表示。

总体指标又称统计指标，简称指标（index），是反映总体特征的数量名称。指标有绝对数指标和相对数指标。绝对数指标又称总量（total）指标或绝对指标，是组成群体的个体标志值总和，反映群体的总量特征。如“少数民族人口数”、“男性人口数”、“老年人口数”、“患病人口数”以及“人年数”等均为群体绝对数指标或总量指标。相对数指标，又称相对（relative）指标或质量指标，是两个有关统计指标之比，反映群体组成结构或相互关系特征。总量与其个体数之比所得的相对数称为（算术）平均数。如“少数民族人口比”、“男性人口比（率）”、“老年人口比”、“患病人口比（率）”、“人口的平均年龄”等均为群体相对指标。其中人口的平均年龄又是平均数。

总量指标又有时期指标和时点指标。时期指标是反映群体在一定时期内形成的总量特征名称。时点指标是反映群体在某一时点总量特征的名称。时期指标与时点指标的区别在前者有累积性而后者无累积性。例如，假定 2001 年某地人口数 100 万，某病发病人数 1 万，则总量指标（人口）100 万是个时点指标，意指该地 2001 年末人口数（不能是全年人口累加）。而总量指标（某病）1 万是个时期指标，意指在该地 2001 年人群中该病总的（累加）发病人数。

总体与个体，因指标与标志由研究目的而定，具有相对性。如研究全国血吸虫病患者在各流行区的分布情况，则总体是全国，（总量）指标是血吸虫病患者数，个体是流行区，（数量）标志是血吸虫病患者。如研究流行区血吸虫病患者分布，则每个流行区又是一个总体，（数量）指标是血吸虫病患者，个体是人，（品质）标志是血吸虫病。在进行综合研究时，由于流行区对人是总体而对全国是个体，所以又称子总体或层。

（四）随机试验、随机事件与随机变量

随机试验是指具有如下 3 个基本性质的观测试验：

（1）试验可在相同条件下重复进行（可重复性）。

（2）试验中所有可能出现的结果明确可知且不止一个（多结果性）。

（3）每次试验前虽不能肯定会出现哪一个结果，但能肯定会出现可知结果中的一个（有结果性）。

随机试验中所有可能出现结果全体组成的集合称为样本空间，组成样本空间的每个可能出现的试验结果称为基本事件。有某种共性的基本事件组成的样本空间的子集称为随机事件，简称事件。试验结果属于某个子集称该子集对应的随机事件发生。如在调查人群年龄结构时，在人群中任意询问一个人的年龄，则询问人的年龄是随机试验，样本空间是非负整数（集），由 0～15、16～45、46～60、61～岁年龄组组成的子集为随机事件。若被询问者年龄为 10 岁，则事件在 0～15 年龄组发生。

试验结果为有限个，每个结果出现的可能性相等的随机试验称为古典试验概型。例如，扔（一颗）骰子试验是个古典试验概型，样本空间为 $\Omega=(i)=\{1, 2, 3, 4, 5, 6\}$；其中每个数出现的可能性都为 1/6。

随机变量是表示随机试验结果即与基本事件对应的变量。随机变量有离散型随机变量和连续型随机变量。离散型随机变量是取值（有限或无限）可列的随机变量。连续型随机变量是可不间断取值的随机变量。如人的心率、细胞数是离散型随机变量，血压、体温是连续型随机变量。

（五）概率、概率函数与数字特征

概率是事件发生可能性大小的数量表示。概率函数是与不同基本事件对应的随机变量取不同值的概率变化，是随机变量的函数。离散型随机变量的概率函数称为概率分布，连续型随机变量的概率函数称为概率密度。随机变量在一定范围内取值的（累积）概率称为（概率）分布函数。随机变量（分布）的数字特征是定量表示随机变量取值整体特征的统计指标。例如，平均值与方差或变异度分别代表随机变量取值（分布）的集中或综合趋势和离散趋势的两个数字特征。

定义在古典试验概型样本空间上的概率称为古典概率。例如，上述扔骰子试验出现点子数 i 的可能性 1/6 是概率，出现奇数点即事件{1，3，5}发生的可能性 3/6 是概率。

随机抽样调查观测具有可重复、多结果、有结果三性，可与随机试验对应。组成被观测群体的个体标志状态与基本事件对应，个体标志（状态）与随机变量（值）对应，与品质标志对应的随机变量为名义变量。与个体标志状态对应的随机变量值组成样本空间。调查总体的统计指标与随机变量的数字特征对应。抽样观测的基本目的是通过样本标志值或随机变量分布特征描述与推断构成总体的个体标志值分布特征，简称随机变量的分布或总体分布。例如，随机试验是从人群中抽样调查以了解人群的性别分布，则由抽取一个人的所有可能性别结果组成的样本空间的基本事件即为男与女两个（标志值）。性别是个无序标志，对应的随机变量是个名义变量，若记男为 0 女为 1，则 0、1 为名义变量值，由所有可能观测结果组成的样本空间的基本事件即为 0 与 1 两个数。通过抽样观测得到的样本（性别）频率可描述与推断该人群总体的性别分布。再如随机试验是从人群中抽样调查以了解年龄分布，若以年计，则由抽取一个人的所有可能结果（基本事件）为非负整数组成样本空间。若以实际时

间计，则样本空间为由代表所有可能结果（基本事件）的正数组成。

（六）寿命分布函数与生存函数

生物个体如人的生存时间称为寿命。寿命是个随机变量，记为 T。寿命分布函数 $F(t)$ 是指个体寿命 T 不超过时间 t 的概率，即

$$F(t) = P(T < t) \tag{1-1}$$

生存函数 $S(t)$ 是指个体寿命 T 超过时间 t 或个体活到时间 t 的概率，即

$$S(t) = P(T \geqslant t) = 1 - F(t) \tag{1-2}$$

（七）随机向量、随机过程与时间序列

随机向量又称多维随机变量，是指由与个体（基本事件）多个不同标志对应的随机变量构成的向量。如若同时考虑人的性别、年龄，则与性别、年龄对应的随机变量构成一个（二维）随机向量。若同时考虑人的性别、年龄、民族、健康水平，则与性别、年龄、民族、健康水平对应的随机变量构成一个（四维）随机向量。

随机过程是指与时间有关的随机变量随时间变化组成的随机变量族。随机过程研究随机变量与时间的函数关系，又称随机函数。随机过程在每个固定时间是个随机变量。

时间序列又称随机序列，是指时间变量只取等间隔整数值的随机过程观测值。例如人的健康水平是个随机过程，每年的生病次数或天数是个时间序列。

（八）统计量、统计分布与统计推断

统计量是样本观测值的函数，也是随机变量。统计量的概率（分布）函数称为统计分布或抽样分布。例如，若记 X_1、X_2、…、X_n 为从以随机变量 X 代表的总体中抽取的容量为 n 的样本，则样本平均数 $\overline{X} = \sum_{i=1}^{n} X_i / n$ 是个统计量，样本平均数 $\overline{X}$ 的分布就是统计分布。又如，若记 x_1、x_2、…、x_n 为样本 X_1、X_2、…、X_n 的观测值，x_1、x_2、…、x_n 按由小到大排列为 $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(i)} \leqslant \cdots \leqslant x_{(n)}$，则取值为 $X_{(i)} = x_{(i)}$ 的随机变量 $X_{(i)}$ 是样本标志值函数，是统计量，称为样本 X_1、X_2、…、X_n 的第 i 个顺序统计量。$X_{(1)}$ 称为最小顺序统计量，$X_{(n)}$ 称为最大顺序统计量。X_1、X_2、…、X_n 的函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)} \\ i/n, & x_{(i)} \leqslant x < x_{(i+1)}, \ i = 1, 2, \cdots, n-1 \\ 1, & x \geqslant x_{(n)} \end{cases} \tag{1-3}$$

表示样本观测值 x_1、x_2、…、x_n 中小于等于 x 的频率分布，称为经验分布函数，也是一个统计量。

统计推断是根据统计量的统计分布对总体（统计）指标或分布作估计或假设检验。估计包括点估计与区间估计，假设检验包括参数检验与非参数检验。估计区间包含总体指标（被估计参数）的概率称为置信概率或信度，记 $1-\alpha$。检验假设被拒绝错误的概率称为犯第Ⅰ类错误（概率），记 α，检验假设被接受错误的概率称为犯第Ⅱ类错误（概率），记 β。

（九）比率、比数与联比

流行病学研究的群体多为有限群体，其中的个体一般具有多标志性。因此，可按个体的不同标志的不同状态，把群体区分成不同构成（composition）部分。如人群除可按民族区分