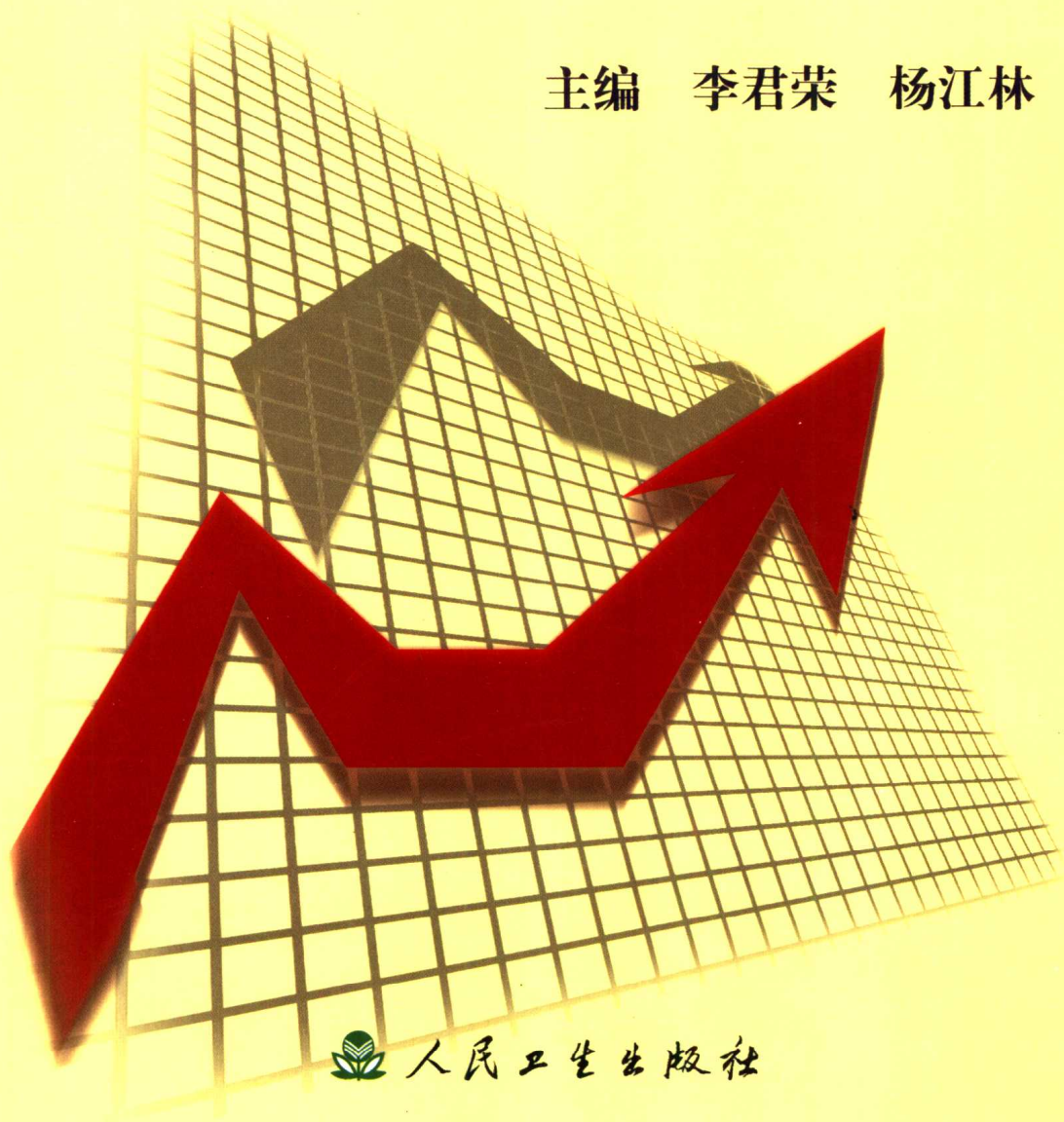


医疗保险专业教材

医疗保险统计学

主编 李君荣 杨江林



人民卫生出版社

医 疗 保 险 专 业 教 材

医疗保险统计学

主 编 李君荣 杨江林

副主编 娄冬华 梁维君 柳晓琳 闵 捷 冯振翼

编 委 (按姓氏笔画为序):

冯振翼 (华北煤炭医学院)	林剑鸣 (广州中医药大学)
吕 媛 (湖南师范大学)	娄冬华 (南京医科大学)
汤质如 (安徽医科大学)	柳晓琳 (锦州医学院)
李君荣 (江 苏 大 学)	赵会仁 (锦州医学院)
杨江林 (郧 阳 医 学 院)	梁维君 (湖南师范大学)
闵 捷 (东 南 大 学)	

人 民 卫 生 出 版 社

图书在版编目 (CIP) 数据

医疗保险统计学/李君荣, 杨江林主编. —北京: 人民卫生出版社, 2003

ISBN 7-117-05344-5

I. 医... II. ①李... ②杨... III. 医疗保险-经济统计 IV. F840.684

中国版本图书馆 CIP 数据核字 (2002) 第 110601 号

医疗保险统计学

主 编: 李君荣 杨江林

出版发行: 人民卫生出版社 (中继线 67616688)

地 址: (100078) 北京市丰台区方庄芳群园 3 区 3 号楼

网 址: <http://www.pmph.com>

E-mail: pmph@pmph.com

印 刷: 潮河印业有限公司

经 销: 新华书店

开 本: 787×1092 1/16 印张: 20.5

字 数: 471 千字

版 次: 2003 年 4 月第 1 版 2003 年 4 月第 1 版第 1 次印刷

标准书号: ISBN 7-117-05344-5/R·5345

定 价: 30.00 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究

(凡属质量问题请与本社发行部联系退换)

医疗保险专业教材
编辑委员会

主任委员 蔡仁华

副主任委员 周绿林 李绍华 李君荣 杨江林

委员 (按姓氏笔画为序)

王 宁	李君荣	李绍华	杨江林	沈华亮
邱鸿钟	官 波	孟国祥	周绿林	张 晓
徐家岳	梁维君	黄占辉	蔡久志	蔡仁华

秘书长 周绿林 (兼)

总 序

中国共产党十六大提出了“全面建设小康社会”的宏伟目标，明确指出，“建立健全同经济发展水平相适应的社会保障体系，是社会稳定和国家长治久安的重要保证。”在医疗保险领域，坚持社会统筹与个人帐户相结合，完善城镇职工基本医疗保险制度；有条件的地方，探索建立农村医疗保险制度。

与此相关，还指出要形成比较完善的全民健身和医疗卫生体系，明显提高全民族的健康素质。这一系列重要指示将引导我国社会医疗保险制度和医疗保健事业的发展与完善，保健康，奔小康。

我国的医疗保险制度历经五十年艰难探索，在上世纪九十年代中后期初步建立了以“基本水平、广泛覆盖、双方负担、统账结合”为原则的城镇职工基本医疗保险模式；农村多种形式的医疗保险制度也在探索与实验中。我们深知任重道远，矢志为建设这项历史伟业而不懈努力！

在社会医疗保险制度改革和发展进程中，我国医疗保险专业教育事业应运而生，逐步发展壮大。近十年来，已有十余所高校相继开设医疗保险专业，为培养我国医疗保险事业所需人才作出积极贡献。为适应新世纪高等院校医疗保险专业人才培养需要，全国医疗保险专业教育委员会精心组织十余所开设医疗保险专业的高校教授、学者和社会医疗保险部门的专家、管理干部共同编写一套医疗保险专业系列教材，既适合作为医疗保险专业学历教育用作基本教材，也可作为广大医疗保险在职工作人员岗位培训用教材或自学参考书。

本系列教材贯彻基本原理、基本知识和基本技能的“三基本”原则，坚持科学性、系统性、启发性、实用性、创新性相结合，充分反映国内外医疗保险理念与体制变革新进展，并密切结合我国医疗保险制度管理与改革的新经验，因而具有较强的学术性和实用性。

我们殷切企盼这套有特色、有创新的医疗保险专业系列教材的出版发行，将有力地促进我国医疗保险管理科学化、规范化，推动我国医疗保险专业教育事业走向新阶段，跨上新平台。

是以为序！

全国医疗保险专业教育委员会

中华医学会医疗保险分会

蔡仁华

2002 年岁末

前 言

随着我国医疗体制改革的不断推进,医疗保险所面临的任务日益艰巨,它不仅要对各种数据进行详尽细致的归纳汇总,进行日常业务反映和业务分析,还要研究如何扩大服务领域,积累保险基金,组织经济补偿,预测医疗服务市场,提高管理水平,制定合理的医疗保险经济发展规划等。医疗保险统计学是进行医疗保险业务分析的有力工具,是探索医疗保险业务活动的特点及其规律性的有效方法。学习和掌握医疗保险统计学的原理和方法,正确搜集、整理和分析医疗卫生服务和医疗保险信息,将有助于我们科学地总结和分析医疗保险现况及其发生、发展规律,为决策部门提供依据,促进医疗保险基金管理、医院管理和卫生经济的发展。

为满足 21 世纪高等院校医疗保险专业人才的培养需要,全国医疗保险专业教材委员会组织目前已开办医疗保险专业的有关院校,充分考虑医疗保险专业培养复合型人才的特点,根据历年在教学、医疗保险基金管理、医院管理实践中所应用的统计方法,总结编著了我国第一本《医疗保险统计学》。本书的撰写内容系统全面、简明实用、适合教学目标,对概念和原理解释力求准确;对理论、方法、技术等知识的介绍,努力做到科学规范;对语言文字的叙述,力戒繁琐。所有统计方法均从实例入手,由浅入深讲述其基本原理、基本概念、基本方法以及各种统计方法的用途和应用条件,便于实际操作时模仿和借鉴,是一本富有特色和创新意义的教材。本书已被列入“江苏省新世纪高等教育改革工程”。

全书共分二十二章,第一~九章、第十三章为基本统计方法,考虑到保险的专业特点,在第二章中增加了总量指标,第七章中增加了动态数列;第十章为线性回归分析(包括简单回归、多元线性回归、逐步回归等),第十一章为相关分析(包括简单相关、复相关、偏相关和典型相关等),第十二章为非参数统计,第十四章调查设计、第十五~十七章为生命统计,第十八章为寿命表,第十九章为 logistic 回归,第二十章为主成分分析,第二十一章为生存分析,第二十二章为时间序列分析等。书中大部分章节附有思考题,可供学生实习、自测和课堂用。

本书既适合于高等院校医疗保险专业及相关专业的本科生和研究生用教材,也可作为学习医疗保险统计方法的广大从事医院信息管理、医疗保险信息管理人员的参考书。

参加本书编写的有汤质如、刘国旗、陈家应、江启成(安徽医科大学),李君荣、刘丽群、唐艳林(江苏大学),杨江林、唐念东(郧阳医学院),闵捷(东南大学),林剑鸣(广州中医药大学),娄冬华(南京医科大学),柳晓琳、赵会仁(锦州医学院),梁维君、吕媛(湖南师范大学),冯振翼(华北煤炭医学院)等。

在本书编写过程中,得到所有参编单位的大力支持,特别是江苏大学、锦州医学院领导的支持。全书的复算由江苏大学医疗保险统计教研室的全体老师完成,特别是刘丽群、唐艳林两位老师所付出的辛勤劳动。特向以上所有对本书给予关心和支持的各位领

导和同仁以及被引用文献的有关作者致以诚挚的谢意。

限于时间和水平，本书不当之处在所难免，我们愿意虚心听取广大读者的批评和建议，以便修订时参考。

李君荣 杨江林

2002.8

目 录

第一章 绪论	1
1.1 统计学发展简史	1
1.2 医疗保险统计学的性质与研究内容	2
1.3 医疗保险统计的意义与任务	3
1.4 学习目的与要求	4
1.5 统计工作的基本步骤	4
1.6 常用的统计学基本概念	5
第二章 数值变量资料的统计描述	8
2.1 频数分布表	8
2.2 集中趋势的描述	10
2.3 离散趋势的描述	15
2.4 总量指标	18
第三章 正态分布及其应用	20
3.1 随机变量的概率分布	20
3.2 随机变量的概率密度和分布函数	21
3.3 正态分布	21
3.4 正态分布在医疗保险中的应用	25
第四章 抽样误差与抽样分布	29
4.1 抽样误差	29
4.2 标准误	31
4.3 常用的抽样分布	32
第五章 统计推断基础与 t 检验	37
5.1 参数估计	37
5.2 假设检验的基本思想与步骤	40
5.3 t 检验与 u 检验	41
5.4 第一类错误与第二类错误	45
5.5 假设检验的注意事项	45
5.6 正态性检验	46

第六章 方差分析	48
6.1 方差分析的基本思想	48
6.2 完全随机设计资料的方差分析	50
6.3 配伍组设计资料的方差分析	53
6.4 多个均数的两两比较	56
6.5 多个实验组与一个对照组的比较	58
6.6 方差齐性检验	59
6.7 变量变换	60
第七章 分类变量资料的统计描述	62
7.1 相对数的概念及用途	62
7.2 相对数的种类	62
7.3 应用相对数时的注意事项	64
7.4 率的标准化法	65
7.5 动态数列在医疗保险中的应用	67
第八章 二项分布与泊松分布	76
8.1 二项分布	76
8.2 泊松分布	80
8.3 二项分布、泊松分布与正态分布之间的关系	82
第九章 卡方检验	84
9.1 两样本率的比较	84
9.2 行×列表资料的 χ^2 检验	87
9.3 配对分类资料的 χ^2 检验	90
9.4 四格表直接计算概率法	91
9.5 频数分布的拟合优度检验	93
第十章 线性回归分析	94
10.1 直线回归	94
10.2 多元线性回归	98
10.3 逐步回归	106
10.4 违背回归模型基本假设的情形	108
10.5 回归分析的注意事项	110
第十一章 相关分析	112
11.1 线性相关	112
11.2 多元线性相关简介	115
11.3 典型相关分析简介	118

11.4 相关分析与回归分析的区别与联系	119
第十二章 常用非参数统计方法	121
12.1 非参数统计概念和应用范围	121
12.2 配对设计的两样本比较——符号秩和检验	122
12.3 成组设计两样本比较的秩和检验	123
12.4 成组设计多个样本比较的秩和检验	125
12.5 等级相关	127
12.6 Ridit 分析	130
第十三章 常用统计图表	134
13.1 统计表	134
13.2 统计图	135
第十四章 调查研究设计与质量控制	141
14.1 调查研究概述	141
14.2 现场调查研究的设计	142
14.3 调查表(问卷)的设计	146
14.4 调查资料的搜集方法	150
14.5 现场调查的质量控制	155
第十五章 静态人口统计	159
15.1 静态人口统计在医疗保险工作中的意义	159
15.2 人口普查	159
15.3 人口性别年龄构成与性比例	161
15.4 人口的地区分布与职业文化构成	163
15.5 人口估计	164
15.6 人口预测	166
第十六章 出生统计与死亡统计	167
16.1 出生统计	167
16.2 死亡统计	172
第十七章 疾病统计	177
17.1 疾病统计在医疗保险实际工作中的意义	177
17.2 疾病统计的对象和观察单位	177
17.3 疾病统计资料的来源	182
17.4 疾病统计常用指标	182
17.5 疾病和死因分类(ICD-10 简介)	184

第十八章 寿命表及其在保险中的应用	189
18.1 人口寿命表概述	189
18.2 寿命表中主要指标和意义	191
18.3 简略寿命表的编制	193
18.4 模型寿命表简介	195
18.5 寿命表分析及其应用注意事项	198
18.6 寿命表在保险中的应用	201
第十九章 多元 logistic 回归	206
19.1 多元 logistic 回归模型的概念及形式	206
19.2 逐步 logistic 回归	209
19.3 条件 logistic 回归	211
19.4 多分类结果变量的 logistic 回归	213
19.5 应用 logistic 回归模型的注意事项	214
第二十章 主成分分析	215
20.1 主成分的定义与性质	215
20.2 主成分的计算	216
20.3 主成分个数的确定及其实际意义	218
20.4 主成分分析的应用	219
第二十一章 生存分析	223
21.1 生存时间及其分布函数	223
21.2 生存资料的非参数统计方法	226
21.3 生存资料分析的参数模型	233
21.4 半参数法	235
第二十二章 时间序列分析	237
22.1 时间序列的概念及符号表示	237
22.2 需求图型与滤波	238
22.3 水平时间序列模型	238
22.4 趋势时间序列模型	240
22.5 季节模型	242
22.6 鲍克斯-詹金斯模型 (Box—Jenkins Model)	244
附 1 统计用表	251
附表 1 标准正态分布曲线下的面积, $\Phi(-u)$ 值	251
附表 2 t 界值表	252
附表 3 χ^2 界值表	253

附表 4	F 界值表 (方差分析用)	254
附表 5	F 界值表 (方差齐性检验用)	258
附表 6	百分率的可信区间	260
附表 7	Poisson 分布 μ 的可信区间	266
附表 8	q 界值表 (Student - Newman - Keuls 法)	267
附表 9	Dunnett 检验临界值表 (双侧)	268
附表 10	T 界值表 (配对比较的符号秩和检验用)	269
附表 11	T 界值表 (两组比较的秩和检验用)	270
附表 12	H 界值表 (三组比较的秩和检验 Kruskal - Wallis 法)	271
附表 13	r 界值表	272
附表 14	Kendall 等级相关系数 r_k 界值表	273
附表 15	r_s 界值表	273
附表 16	随机排列表 ($n=20$)	274
附表 17	随机数字表	275
附 2	思考与练习	277
第一单元	绪论	277
第二单元	数值变量资料的统计描述	277
第三单元	正态分布及其应用	278
第四单元	抽样误差与抽样分布	280
第五单元	统计推断基础与 t 检验、方差分析	280
第六单元	分类变量资料的统计描述	284
第七单元	二项分布与泊松分布	287
第八单元	分类资料的统计推断	288
第九单元	线性回归分析和相关分析	291
第十单元	多元线性回归	293
第十一单元	非参数统计	295
第十二单元	统计图表	297
第十三单元	生命统计	299
附 3	常用英汉统计学词汇	303
附 4	参考文献	314

第一章 绪 论

当今社会是信息时代,信息的传播与交流日益加快。人们如何去分析、挖掘信息,达到占有信息的目的,就要对事物的变化有更深入的认识。事物的变化现象有必然和偶然之分,且经常是偶然性(不确定性)掩盖了必然性,妨碍了人们对客观规律的认识。如何去伪存真、去粗取精,正确地 from 表面纷繁杂乱的不确定性现象中揭示其内在的规律,正是统计学(statistics)所面临的根本任务。

1.1 统计学发展简史

统计一词由来已久,它的涵义在历史上不断演变。拉丁语 state 的意思是各种现象的状态和状况。在原始社会里,人类最初的一般计数活动,就蕴藏着统计萌芽。随着奴隶制国家的产生,统治阶级为了对内统治和对外战争的需要,需要征兵征税,开始了人口、土地和财产及世袭领地的统计。进入封建社会,统计已略具规模,一些君主和开明的政客日益认识到统计对于治国强邦的重要作用。到了封建社会末期,特别是进入资本主义社会后,社会生产力迅速发展,社会分工愈益精细,交通、航运、贸易日益发达,国际市场逐步形成,政府不仅需要有人口、土地、财富、赋税和军事等国情国力方面的统计,各企业主、商人为了经营管理也需要统计,统计逐步扩展到了工业、农业、商业、银行、保险、邮电、海关等部门,并出现了专业性统计机构和研究组织。在这样的历史条件下,17世纪中叶以后,随着统计实践的发展,客观上要求总结经验,使之上升为理论,并进一步指导实践。欧洲开始出现一些统计理论著作,统计学(statistics)应运而生,并出现了一些不同的学派。

1.1.1 古典统计学时期

这是统计学的萌芽期,指17世纪中叶至18世纪中叶。主要有两大学派。

(1) 政治算术学派:17世纪中叶英国,代表人物是 W. Petty,代表作是《政治算术》。在这部著作中,首次用计量和比较的方法对英、法、荷三国的经济实力、军事实力和内在潜力进行了分析,他主张一切论述都用数字、重量和尺度来进行,并提出用图表形式概括数字资料的理论和方法。其理论与方法对后来的统计学形成与发展有深远的影响。

(2) 国势学派(亦称记述学派):该学派产生于18世纪的德国,主要代表人物是 H. Coning 和 G. Achenwall。Achenwall 一个很主要的贡献是,他第一个把国势学定义为“statistik”,统计学的名称一直沿用下来。该学派后来逐渐发展为政府统计。

1.1.2 近代统计学时期

这是统计学的发展期,指18世纪末至19世纪末。在这段时期统计学发展并形成了

许多学派,其中最主要的是数理统计学派和社会统计学派。

(1) 数理统计学派:它是在概率论已有相当发展的基础上,将概率论引进统计学而形成的。其奠基人是比利时学者 A. Quetelet, 他的老师 P. S. Laplace 主张用自然科学的方法研究社会现象,并正式将概率论引入统计学,使统计学进入了一个新的阶段。他最早用大数定律论证了社会生活现象复杂变化的偶然性中存在着规律性,并提出了误差理论,用以解决统计学的准确性问题。并将统计学明确定义为一门既研究社会现象,又研究自然现象的独立的方法论科学,在英美等国得以发展,至 19 世纪末,正式从统计中分离出来,自成一派。

(2) 社会统计学派:该学派是以德国为中心, K. G. A. Knies 首创,主要代表人物有 C. L. E. Engel 和 G. V. Mayr 等人。他们认为统计学是一门社会科学,是研究社会现象变动原因和规律性的实质性科学,以此与数理统计学派的通用方法论相对立。社会统计学派研究社会总体而不是个别的社会现象,将政府统计与社会调查结合起来,形成自己的特点,在国际统计学界占有一定的地位,对日本等国的统计学界有一定的影响。

然而,社会经济不断发展,要求统计学提供更多的方法,社会科学本身也不断向量化发展,日益重视方法论的研究,出现了从实质性科学向方法论科学转化的趋势。

1.1.3 现代统计学时期

是指 20 世纪初迄至今日的统计学发展时期。这一时期的科学技术迅猛发展,社会生产发生巨大变化。数理统计学同自然科学、工程技术紧密结合而被广泛应用,获得迅速发展,进入了鼎盛时期。首先, W. S. Gosset 提出了小样本的 t 分布理论,其后, R. A. Fisher 加以充实,并由 J. Neyman 等人进一步发展,建立了统计假设理论。后来, A. Wald 创立了决策理论, S. S. Wilks 和 J. Wishart 等统计学家对样本分布理论加以充实和发展, R. A. Fisher 首次提出了实验设计的概念和三个基本原则(对照原则、随机化原则和重复原则), W. G. Cochran 等对其丰富和发展,进一步拓宽了统计学的范围。在这期间,特别要提出的是数理统计学派吸取了生物学的有益成果, F. Galton, K. Pearson, W. S. Gosset, R. A. Fisher 等一批杰出的统计学家,提出并发展了回归分析、相关分析、假设检验等方法。20 世纪 60 年代后,数理统计学的发展呈现出三个明显的趋势:一是越来越广泛地应用数学方法;二是出现了一些新的分支(如抽样理论、非参数统计、多元分析、时间序列分析等)以及一些边缘学科(如经济计量学、医学统计学、天文统计学、工程统计学等);三是应用更加广泛,尤其伴随着计算机的发展,各种统计软件层出不穷,使其所发挥的作用也愈加明显。因而,数理统计学派已成为现代统计学的主流学派。

在各学派争论的同时,出现了一支不可忽视的学派,就是 Bayes 学派。该学派在近 30 年中逐渐形成体系,也开始走进教室,开始打破数理统计学派一统天下的局面,值得研究人员关注。

1.2 医疗保险统计学的性质与研究内容

统计学原理和方法几乎应用到自然科学和社会科学的各个领域,产生了许多应用性

分支, 诸如社会经济统计学、工业统计学、生物统计学、卫生统计学、农业统计学等。如果将统计学原理和方法应用到医疗保险领域, 由于该领域的研究对象是人, 具有其特殊性, 并受诸多的社会、经济和心理因素等的影响, 许多现象均具有不确定性, 统计学更是大有用武之地, 完全有理由成为一个统计学的分支。

医疗保险统计学 (Statistics in health insurance) 是运用概率论和数理统计原理、方法结合医学、保险学的实际情况, 研究医疗保险现场调查的设计以及数字资料的搜集、整理、分析和推断的一门科学, 是进行医疗保险日常管理和科学研究所必需的重要手段。它不仅是一门方法学, 同时也有自身研究领域的一门科学。例如, 在保险业务中要检查计划完成情况, 动态分析保险业务收入情况, 计算各种费额, 对社会经济现象和年保费增长趋势进行预测, 以及填报、计算各种统计报表, 有时为了预测保险给付额, 常常要分析疾病登记资料, 了解疾病的发生发展规律, 这些都离不开统计学。医疗保险专业的学生学习统计学, 其目的不言而喻。通过本学科的学习, 要求同学充分认识原始资料的重要性, 如实反映研究中的现象是每个科学人员必备的品质, 记住: “Figures don't lie but liars do figures”。

医疗保险统计学的研究内容非常广泛, 几乎所有的统计原理和方法都能在医疗保险领域得以直接或间接的应用, 根据目前该领域研究的现状以及教学学时的限制, 本书着重介绍的主要内容有: 调查设计, 统计分布理论, 样本资料的统计描述, 统计推断 (包括参数估计及假设检验), 相关分析, 回归分析, 生命统计, 多因素分析, 生存分析以及统计预测等。

1.3 医疗保险统计的意义与任务

1.3.1 医疗保险统计的意义

医疗保险产生的自然条件是生死病残等事件的客观存在, 其经济基础是商品生产和商品交换, 统计则是医疗保险经营管理和发展的技术手段。病死伤残事件的发生时间、对象和程度是无法确定, 但可以运用统计学原理和方法, 揭示这些现象变化的关系和规律。如医疗保险费率的确定, 是根据大量的生命统计资料以及其他的有关因素计算的。充分认识医疗保险统计的意义和作用, 对于全面了解医疗保险的现状, 制定医疗保险政策、改善医疗保险经营以及促进医疗保险事业的发展有着很重要的现实意义。

1.3.2 医疗保险统计的任务

(1) 为国家或地方制定医疗保险工作的方针、政策提供依据: 医疗保险大政方针的制定是建立在有关医疗保险、经济活动、卫生服务水平、居民健康状况等分析的基础上, 不是“一拍脑袋”形成的。

(2) 分析评价医疗保险业务经营情况和影响因素: 医疗保险工作的好坏, 无非从其科学性、先进性和效益等方面进行评价。其中不乏有大量的统计指标和指标体系, 甚至多因素统计分析, 如主成分分析用于医疗保险评价。

(3) 分析疾病的发生、发展规律：医疗保险的经营、管理及科学研究都与疾病的发生与发展分不开，需要经常了解居民的健康水平，分析疾病的发生、发展规律，研究医疗费用的结构、消长，进而调整医疗保险费率、降低管理费用。

(4) 建立、健全保险统计档案：对于日常管理以及不同时期的比较，甚至不同地区间的比较积累资料。

1.4 学习目的与要求

1.4.1 学习目的

本课程的学习，并非让每位读者成为统计学家。学习本课程的目的在于使学生掌握统计学思维方法，学会从不确定性和概率的角度去考虑问题。一要能够结合自己所学的专业知识去合理地设计调查，获得可靠、准确的资料，充分挖掘资料中隐含的信息，恰如其分地作出理性的概括，写成具有一定水平的调查报告或科学论文；二要能运用自己的统计学知识去批判性地阅读、评价已发表的文献。这些对于提高自身的科研素质有非常积极的意义。

1.4.2 如何学习

由于本学科具有抽象性和实用性，因此学习方法不同于一般的学科，作为一个应用工作者更讲究实用性，因此，对于书中的公式来源不必深究（有兴趣的同学可自己查阅有关书），只要求读者重点掌握以下几点：

- (1) 基本概念；
- (2) 统计计算方法及其应用条件；
- (3) 结合专业知识，对统计结果作出正确的解释。

为提高理论教学效果，学员应重视实习课，以加深对统计理论的理解和培养自己解决实际问题的能力。

1.5 统计工作的基本步骤

统计工作的步骤类似于科研工作的步骤，一般分为四个步骤：设计（design）、搜集资料（collection of data）、整理资料（sorting data）、分析资料（analysis of data）。前一步骤是后一步骤的基础，后者又对前一步骤提出要求，相辅相成，缺一不可。任何一个步骤的缺陷，都会影响统计分析的结果。

(1) 设计：统计工作首先要有一个全过程的设计，可以说这是很重要的一步（详见第14章）。

(2) 搜集资料：资料搜集的好坏是统计工作（科研工作）的关键，如果搜集原始资料的计划不周密，原始记录不准确，就会造成整理资料、分析资料的困难，甚至得出错误的结论。

资料的来源可以有多种渠道。①各医疗部门、医疗保险网点的统计报表；②日常医

疗卫生工作记录和报告卡。这些属于经常性资料。为了使这些资料可以供医学科研使用, 必须注意资料的完整性、及时性和准确性, 防止重复和遗漏或者项目填写不清等情况发生, 一旦发现错误, 应及时纠正, 否则收集来的资料也不可靠。因此, 医疗卫生人员必须从思想上认识原始记录的重要性。③专题调查和实验性研究资料。有些资料不能通过上述途径得到, 如百姓对现行医疗保险政策的看法, 某种政策对医疗费用结构的影响等等, 必须进行专题调查或实验才能取得所需的资料 (详见第 14 章)。

(3) 整理资料: 又称为统计归纳。即根据研究设计者整理分析计划的要求将资料进行分组与汇总, 使其条理化、系统化, 以便分析。需要说明的是, 在整理资料之前, 务必核校原始资料, 一旦发现可疑之处, 立即采取措施加以复核和纠正。

(4) 分析资料: 是对整理过的资料作进一步的统计计算、分析的过程。统计分析包括统计描述和统计推断, 其结果可用适当的统计图表等表示。同时, 还必须结合专业知识给出恰如其分的专业结论。

1.6 常用的统计学基本概念

1.6.1 同质与变异

同质是指性质相同或相近的事物。统计学常要求的观察对象同质, 是指性质相同或相近的观察对象。例如, 研究某种疾病, 一定要有统一的诊断标准; 研究儿童身体发育情况, 同性别、同年龄的儿童称为同质观察单位。然而, 即使是同质观察单位, 就其某一特征 (观察指标) 而言, 个体之间也是存在差异的, 这种差异称为变异 (variation)。例如, 同年龄同性别的儿童, 其体重不一定相同; 某药物治疗一疾病, 有的患者能治愈, 有的患者好转, 有的患者则无效甚至恶化、死亡。

1.6.2 总体与样本

总体 (population) 是指根据研究目的所确定的同质研究对象的全体, 它是由客观存在的具有相同或类似性质的研究对象组成。例如, 我们要调查镇江市在职职工的收入情况, 该市所有的单位或企业在职职工就构成了本次研究的总体, 其同质的基础是工作在镇江市范围内的单位或企业在职职工。样本 (sample) 是从总体中随机抽取的部分有代表性的观察单位, 某指标的实测值即构成了样本。如上, 若了解镇江市在职职工的收入, 可以对总体中的所有职工进行调查, 称之为普查, 但限于时间、人力、物力和财力等方面的影响, 普查的难度比较大, 如能达到研究目的, 甚至是没有必要的。此时, 可以在总体中随机抽取一部分有代表性的单位或企业职工作为本次研究的调查对象, 该部分职工就组成了本次研究的样本。

1.6.3 参数与统计量

描述总体特征的统计指标统称为参数 (parameter), 用希腊字母表示。而由样本的信息所计算的统计指标, 称之为统计量 (statistic)。例如, 镇江市所有职工的平均收入就是一个参数, 而根据职工样本所计算出的平均收入则是一个统计量。