

当代保险译丛

〔英〕 G. C. A. 迪克森 著

应用保险统计学

DANGDAI BAOXIAN YICONG

陈宪平 阎晓田 译 王育宪 校



《当代保险译丛》出版说明

英国特许保险学会(CII)为培养高级保险专业人才,每年对保险从业人员进行资格考试,历史已久。考试分初、高两级,对初级考试合格者,学会授予准会员(ACII)资格;对高级考试合格者,学会授予正式会员(FCII)资格。这种资格是英国保险专业的最高学位,也为国际保险界所公认,世界许多国家都有它的考试代办机构。

中国保险学会为引进国际保险界所公认的保险专业知识,发展国际保险业务,于1986年成立了代办英国特许保险学会的考试中心。现取得该学会的同意,将该学会考试专用的一系列教材,有选择地翻译,编成《当代保险译丛》陆续出版。以普及国际保险的专业知识,并为我国自办保险干部资格考试以及编撰我国自己的保险教材打下基础。

英国是世界上保险历史比较悠久的国家之一,保险业务比较发达,伦敦市场的保险业务活动对国际保险界有较大的影响,同时有关保险理论和实务的著作也比较成熟,因此,我们首先翻译英国特许保险学会的专用考试教材,作为普及保险知识的开端。第一批拟出版20种:其中有《保险入门》、《法律原理》、《保险原理与实务》、《保险经济学》、《养老金和有关利益》、《个人保险》、《财产和利益保险》、《责任保险》、《人寿保险法和实务》、《海上保险法》、《航空法》、《海上保险承保》、《航空保险承保》、《海上保险理赔》、《航空保险理赔》、《汽车保险》、《应用保险统计学》、《财产保险的风险估计与控制》、《再保险原理》、《保险公司管理》。

保险是一门社会科学，同各国的社会制度和法律密切相关。因此，在阅读这些著作时应了解原作者国家的社会与历史背景，加以分析取舍；在应用时，要结合我国国情，不宜照搬。

本《译丛》由一部分多年从事保险工作的专业人员以及从事经济学研究或教学的学者翻译，并请有关专家审校。但因各人对原著理解难免有误，因此对译文中的不足之处，欢迎广大读者批评指正，以便修订。

本书承黄河大学经营管理系主任李中杰教授审阅，在此表示感谢。

中国保险学会
《当代保险译丛》编辑委员会

前 言

在本书编写过程中，由于充分考虑了许多学生在学习保险统计学时的困难，因此，本书则以深入浅出地介绍保险统计学的知识为宗旨，以期最终使读者有信心胜任保险业中的数量分析工作，并领会到学习本课程的意义。

本书的读者可分为两类。一类是具有保险数量分析知识的读者，他们多半是已在大学里学过某些统计学或数学课程的毕业生，包括某些熟悉统计学以及毕业于数学专业的学生，这部分读者可按章节的顺序学习本课程。另一类则是原来不懂或缺乏统计分析知识的读者，这类读者可能是在校没有学过数学或没有学好数学的毕业生，以及更多的是那些已毕业多年并早已将在校学习的数学知识忘掉的一部分学生。因此，这类读者必须首先学习本书的最后一章，补上所要求的基本的算术和代数知识；然后再从第一章起，顺序地学下去。

其实，学好保险统计学并不神秘，关键在于多作练习。这样可免除许多学生在初学本课程时产生多余的忧虑。应该指出，大多数的统计学的初级教材中都包含了大量的习题。读者应尽可能多读一些课外材料，以便熟练地掌握各种统计和定量分析的技巧。

本书各章给出了许多保险实例。其中在第三章结尾部分列举的实例，综合了前三章所学的技巧。同一例题再次用在第六章的末尾，以此来应用所学过的全部内容。

本书的最后附有符号与公式速查表。

目 录

第一章 数据收集	1
1.1 统计学的特征与含义	1
1.2 总体和样本	4
1.3 数据的收集方法	6
1.4 随机样本	9
1.5 非随机样本	11
1.6 分层抽样	13
第二章 数据表示	15
2.1 离散型数据和连续型数据	15
2.2 简单数组	16
2.3 频数分布	18
2.4 条形图	27
2.5 饼形图	30
2.6 象形图	31
2.7 散点图	32
第三章 数据描述	33
3.1 数据描述的必要性	33
3.2 定位的度量	34
3.3 离差的度量	50
3.4 偏斜度的度量	58
3.5 第一至三章知识应用实例	61
第四章 概率与风险理论	71
4.1 概率的性质	71

4.2 概率的求解	72
4.3 概率的应用	75
4.4 概率分布	80
4.5 风险理论	91
第五章 统计抽样	97
5.1 抽样的重要性	97
5.2 总体均值的估计	98
5.3 总体比例的估计	112
第六章 假设检验	117
6.1 假设总体均值的检验	118
6.2 假设总体比例的检验	129
6.3 样本均值差的检验	131
6.4 样本比例差的检验	133
6.5 第四至第六章知识应用实例	135
第七章 回归与相关	140
7.1 变量间的关系	140
7.2 简单线性回归	145
7.3 相关	152
7.4 应用实例	165
第八章 指数与时间序列	169
8.1 指数	169
8.2 时间序列	187
第九章 精算的概念、原则与术语	200
9.1 风险	200
9.2 保险费	204
9.3 保险单	219
9.4 保险业务量	222
第十章 数据处理系统	224

10.1	数据处理的发展过程.....	224
10.2	电子计算机.....	235
10.3	数据处理.....	264
10.4	远程处理.....	277
10.5	应用实例.....	278
第十一章 统计学的数学基础		862
11.1	定义.....	286
11.2	算术运算.....	287
11.3	代数方法.....	295
11.4	精确与近似.....	303
符号与公式速查表		308

第一章 数据收集

1.1 统计学的特征与含义

在过去的数十年里，定量法或通常所说的统计方法有了极为显著的发展。我们几乎每日都要接触超出我们的信息处理能力、却又为我们必须了解的统计数据。报刊、杂志、广播和电视也都心甘情愿地成了统计学的主持人。人们由此可能得出这样的结论，统计学出现于最近这些年。然而事实并非如此。历史地看，统计学不过是一门描述性的算术，专门用来为经济学家和某些人提供关于社会情况的更精确的资料。在早期使用该方法的人们中，有一位埃德蒙德·哈雷 (Edmond Halley) 先生(无人不晓的哈雷彗星便因他而得名)。为确定保险费，他制作了生命表。今天，统计学已成为一门关于收集和整理数据资料，进而分析并从中得出结论的科学。前者称为描述统计学，后者称为分析统计学。

1.1.1 统计学的应用

实际上，统计学的运用范围极其广泛。政府运用统计学来制定政策；军队应用统计学改进后勤；药品工业应用统计学推动科研；商业上应用统计学完善计划。以上仅是统计科

学在多种领域中所显示其功能的几个事例。统计学的运用无疑给世界带来了巨大的变化，这种变化在商业环境中尤为明显。统计学在商业中的运用已经使经营管理、金融、销售、广告宣传以及许多其他社会领域发生了变化。

为说明这一点，让我们来考虑一个兼营人寿保险与非人寿保险的保险公司所极为关注的一些问题。为了预测现有险种和新险种的销路，总体趋势的预测是极为重要的；成本预测则是使用新的工作系统（如微型信息处理系统）或调整人事安排的基础；在设计和签发新保单以及确定投资策略时，预期利润的估算又是必不可少的；此外，办公室的布局、大小、地理位置和维修保养，缮制保单时的差错率和质量控制，业务成绩分析和预算分配等等，均须确定。上述诸方面都涉及到决策问题。毫无疑问，在这些场合，现代统计学是大有可为的。

1.1.2 统计学的含义

统计信息的价值是显而易见的。但是，真的如此吗？迪斯雷利(Disraeli)曾有过这样的论述：“有三种谎言，即一般的谎言、令人诅咒的谎言和统计数字。”确实，当我们听到平均每一个家庭有2.5个孩子的统计数字时，我们便要联想到这一表述的事实本身。有谁曾见到过0.5个孩子呢？

然而，迪斯雷利仅反映了这一事实，即在许多情况下，统计学有可能不会得到公众的广泛支持。当听到“某些数字不是谎言”或“它们不会产生争议”时，我们常会对此产生疑问吗？当然，统计分析的计算可能是相当精确，然而计

算所依据的问题的基本性质是什么呢?如果基本性质在某些方面是靠不住的,那么最终结果很可能是完全错误的结论。这就如同在烘制一个相当精美的草莓糕饼时,虽然所有原料的比例恰当,但主要原料用的不是面粉而是沙子的情况。

可以说,统计学的阐述包括两个组成部分。第一,被分析对象的基本性质;第二,以数学和图表表示的最终结果。本书绝大部分内容致力于阐述统计学的后一部分内容,但是在论及该点之前,我们有必要了解正在考虑的分析对象的基本特征的两个方面。

1.1.2.1 有偏资料来源

在获取所需资料信息时,你可能会遇到具有偏差的资料。对此,必须将这种偏差限制在最小范围内,甚至将其消除。

如果有人询问你目前的薪金,你是否愿意分文不差地告诉对方?或许你倾向于言过其实?或许稍有保留?很明显,对于统计学家来说,他必须断定上述情况是不存在的。这要求他仔细考虑谁在回答问题,为什么要这样回答等等。他必须判断信息来源的有效性以及真实性。

假定你向投保汽车险的全部保户发出一张调查表,了解他们对不附带无赔款退费条件的汽车险新保单是否感兴趣。根据收回的调查表可以断定,80%的保单持有者对此是感兴趣的。但是如果仅有10%的保单持有者回答问题,那么又该如何下结论呢?

再举一例,某保险分公司的经理,建议将办公地点迁至乡村,于是人们则十分关心使用定期车票的效果,同时,也担心那些使用定期车票的人员是否过分地埋怨英国的交通政

策。这里，主观偏差是存在的。但何谓过分？埋怨由何产生？什么是交通？政策应如何解释？弄清这些概念是十分必要的。

1.1.2.2 证据不充分的论点

在另一种情况下，可能所得到的信息资料并不存在偏差，但是却在此基础上做了不合逻辑的演绎。这种情况是如此危险，以至于统计学家必须确保自己的论据无懈可击。为此，人们总是要对同一现象找出几种可供选择的解释。

试考虑如下论述：

“斯特灵(Stirling)的死亡率低于其他任何城市。因此，斯特灵是全国健康水平最高的城市。”

在找到其他答案之前，也许你能够接受上述论点。但是在寻找其他解释时，当你发现所有在斯特灵发生的死亡，没有在当地登记，却跑到格拉斯哥(Glasgow)登记，你的想法就会不同了。

我们生活在现代的社会里，面对“过去50年里私生子增长了一倍”这一表述，其结论如何？一方面，确有更多的私生子出生的记载，但是另一方面，新生儿死亡率也已大大降低。人们可能会各执一端，争执不下。

由此可见，必须细心谨慎，并提倡善问的方法和运用逻辑思维。这些均是应用统计方法最重要的几个问题。

1.2 总体和样本

在学习统计学时，有必要分清总体和样本的区别。如果一组数据由可以反映一定现象的全部观察值组成，即可定义为总体。例如全体雇员、全部代理帐户、全部地方分支机构

均可称为总体。如果数据仅包含其中一部分，则称为样本。

现以一例说明。一家公司的某部门拥有50名理赔人员。公司经理认为人员可能过剩，因此为了提高该部门的劳动生产率，必须对人员进行一些有效的分析。如果从全部50名职员中抽取10名，并以这10人作为统计计算的基础，这10人便构成一样本，而全部50人则为总体。若换个角度，更高层次的管理可能需要在全国基础上做一类似的分析，50名职员此时便成为全国范围意义上的一个大总体的样本。总体一词是统计学术语，因为在英文中，它和人口学中的人口共用一词，但却与人口无任何关系，而更多地与实际（或假设）的计量的全体有关。

实际上，统计学中的样本一词也比其原来范围较广的常用含义要狭义的多。例如，该词原意常常用来表示一种尝试或试验（如香烟样品或一辆新汽车的试车）。在统计学中，该词便与上述含义大不一样，正如我们将要看到的，它仅用来表示一些数据，通过这些源于总体的数据，即可对总体进行无偏差且有效的分析。

十分清楚，为了获取信息资料，分组的含义必须清晰准确。一家综合性保险公司的男性保单持有者与另一家综合性保险公司分公司的男性保单持有者是完全不同的总体概念。后者主要取决于分公司所在的地区，因为它将反过来影响收入、社会等级以及雇佣性质等诸多因素。

1.2.1 定性定量

在进入讨论数据收集方法之前，最后需要区分一下数据

的定性与定量特征。定性特征涉及了相对于变量数量而言的品质，当根据性别对办公室的全部人员进行分类时，这就是定性。另一方面，当作关于全部雇员人数的表述时，这就是定量方法。

1.3 数据的收集方法

任何统计分析必须取得数据，然而这里所指的数据必须是分析的对象所必需的。收集与之无关的数据则是一种时间、精力以及财力上的浪费。如果所需统计资料仍未得到，那么，调查者必须决定应采取何种收集方法。以下将介绍五种方法：

- (1) 提取法；
- (2) 观察法；
- (3) 走访法；
- (4) 实验法；
- (5) 调查法。

1.3.1 提取法

保险人常常依靠政府的统计资料来确定他们今后的策略。政府统计部门（中央统计局）一般公布如下资料：

- (1) 年度统计摘要（包括反映各类事实的数字，如气象、教育、劳动力、商品零售、国家财政、保险、银行业等等）；
- (2) 统计月报（该月报与（1）极为相似，但不含银行与保险的统计数字）；

(3) 经济动向(给出表明经济现状的统计数据);

(4) 社会动向(提供国家的社会条件的数据)。

上述全部例子都是第二手数据。换句话说,它们是其他一些人的研究成果。调查人员通常不晓得这些数据的背景来源,因此,在使用这些数据时必须慎重。如果这类数据仍无使用价值或无处索取,则研究人员只得依靠各种手段自己动手去收集数据。这样的数据是第一手资料来源,因为研究人员自己收集数据,能够意识到数据来源的条件以及数据本身的局限。

1.3.2 观察法

可以论证,直接观察法是最好的方法,因为它使收集错误数据的可能性降至最小限度。然而直接观察法并非在任何场合都适用,例如,如果你准备调查某些保单持有人为什么总采取特定的行为方式,那么若使用直接观察法或许没有多少用处;但是,如果你想得出某一年份间所售出的保单数量,那么最好的方法是去数它们。显然,任何一种方法的适用性取决于调查对象的性质。

1.3.3 走访法

在诸种收集数据的方法中,走访法是较好的方法之一。但是,除非走访者本人受过良好的训练,否则,采用这种方法去搜集资料的代价是很高的。若走访者未经良好训练,则极可能出现偏差,而由此导致大量严重的资料来源差错。虽然走访在经济上不合算,但如果使用得当,此方法能够克服

采用调查表出现的反应率低的问题。然而，走访方式也存在以下一些问题。如果被访者不再复述一遍，你是否还能记住前面所谈的一段话最开头的四个字？这恐怕难以做到。健忘是这一手段的主要缺点之一，走访者如果不将采访的每一件小事都记录下来，在某些时候就必须依靠其记忆力去记录他走访时的全部发现。

此外，走访方式还有其他一些不利之处。例如对下述提问：

“你驾驶汽车横穿过某城市几次？”

回答者可能会还为提问者仅仅关心他什么时间驾车而误解了该问题的内容。的确，如果他感到如实回答会造成其汽车保险费的相应增加，他甚至可以编造一个谎言。

1.3.4 实 验 法

如果保险人想去检验一套新的自喷淋系统的效果或新防盗报警装置的价值如何，在其客户的建筑物内暗中安排一次系统装置的运转是不妥当的。测量设备或装置效果如何的恰当方法是实验法。

例如，一家大保险公司的市场部估计一张新设计的保单、新的广告策略或市场费率等对消费者购买保险的影响，则实验法是收集适用数据的理想方式。

1.3.5 调 查 表

调查表也许是最经济的方法，可用于从数量众多的人们中(如分支机构的经理们)收集数据。但是，该方式在许多方

面仍是弊大于利。

(1) 调查内容常被误解;

(2) 反应率低;

(3) 对调查表作出答复的人可能有某种既得利益。

在设计调查表时, 下述各点必须牢记:

(1) 调查表以及类似的问题应简明扼要, 清楚易懂。不能模棱两可, 含混不清。

(2) 不应使用诱导性的问法。

(3) 应选择能作出简要回答的问题, 选择答案方式比要求对方作出文字回答的方式更为优越。选择答案方式也可用于简单的问题分析。

(4) 应避免提及个人性格之类的问题。

以下是一个这类不宜提倡的提问形式的例句。

“你不同意所有未向非统一费率公司购买保险, 却提出索赔的汽车驾驶员们不应享受其无赔款时的减费奖励吗?”

1.4 随机样本

为获取数据, 研究人员首先须明确地限定其总体性质, 随后确定抽样结构。简单地讲, 抽样结构即为一张反映其全部总体单元的表格, 这对于产生一个随机样本将十分必要。

例如, 拟对人寿险保单持有者中的全部男性进行调查, 依据保单编号按数码顺序建立全部记录的抽样结构自然将女性保户也包括在内, 这是不合适的。而按性别分别重新排列将建立一个正确的抽样结构, 使之能够得到随机样本。

1.4.1 偏 倚

那么，什么是随机样本？假定在所建立的抽样结构中，有400张标明男性的卡片，依顺序拿出前40张作为样本，这就会产生偏倚现象。例如，可能前40张卡均为期限较长的保单，这是因为卡片是依保单编号顺序排列的，可能时间长的排在前面。随机样本所要求的是在选择40张卡片时，要保证400中的每一张都有被选取的同等机会。

这就是随机样本的确切含义。但要注意，随机抽样作为选择手段，仅仅在选择方法上消除了偏倚，但不能消除来自样本本身的偏倚，尽管在某种程度上能减少这种偏倚。即使进行随机选择，仍可能抽到年龄超过75岁的男性样本，由此而产生一偏倚样本，尽管该样本同样是以随机方式收集而来。

1.4.2 随机数的产生

至少有以下三种选取随机样本的方法。

(1) 可将全部编有序号的卡片扔进一个桶内，并反复搅之，从而使我们所需要的号码混入其中。由于显而易见的原因，此方法是不足取的。

(2) 拟随机样本。现有400张卡片，我们需要其中40张，因此每10张中选取1张，同时要求被选取的这张必须从10张卡片中随机抽取。此方式的不足之处在于存在着重复的方式，因而会由此产生偏倚。例如，可能每10张卡中都有30年期的卡片，有时人们称这种抽样方法为系统抽样。

(3) 最后一种方法即使用特制表格以得到随机数，或在