

世纪保险精算系列教材

精算师考试用书

中国人民大学风险管理与精算中心主编

精算中常用的 统计模型

黄向阳 金阳 肖宇谷 宁威 编著

 中国人民大学出版社

21 世纪保险精算系列教材

精算师考试用书

中国人民大学风险管理与精算中心主编

精算中常用的统计模型

黄向阳 金 阳 肖宇谷 宁 威 编著

中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

精算中常用的统计模型/黄向阳等编著.

北京:中国人民大学出版社,2009

(21世纪保险精算系列教材)

ISBN 978-7-300-10255-9

I. 精…

II. 黄…

III. 保险-精算学-统计模型-教材

IV. F840.4

中国版本图书馆 CIP 数据核字 (2009) 第 007057 号

21 世纪保险精算系列教材

精算师考试用书

中国人民大学风险管理与精算中心主编

精算中常用的统计模型

黄向阳 金 阳 肖宇谷 宁 威 编著

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511398 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn		
	http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京雅艺彩印有限公司		
规 格	170 mm×228 mm 16 开本	版 次	2009 年 2 月第 1 版
印 张	23	印 次	2009 年 2 月第 1 次印刷
字 数	416 000	定 价	32.00 元

版权所有 侵权必究

印装差错 负责调换

总 序

自 1775 年英国公平人寿最早将运用数学工具为产品定价的专门人员命名为精算师以来，精算师职业在国际上已有 200 多年的历史。这一职业最早在人寿和养老金业务中发挥作用，之后逐步向非寿险、健康保险、社会保障等领域扩展。20 世纪以后，精算师的职业进一步延伸到银行、投资、公司财务、金融工程等领域。精算师职业领域的扩展与精算职业组织的发展和精算教育水平的提高密切相关。1848 年后欧美一些国家陆续成立的精算师协会以及国际精算师协会，为提高全球精算教育标准作出了贡献。例如，国际精算师协会早在 1998 年就公布了初级精算教育标准，要求 2005 年后加入国际精算师协会的成员在精算教育标准上符合国际教育标准。2007 年，国际精算师协会再次公布了重新修订的初级精算教育标准及教育大纲。国际上著名的精算师职业组织，包括北美寿险精算师协会、北美非寿险精算师协会、英国精算师协会等，也从 2000 年后陆续对其精算教育标准和精算师考试体系进行改革，强调精算学与统计学、金融学、投资学、会计学、经济学等学科的融合，强调精算学科培养复合型风险管理人才的目标。

我国精算教育和精算师职业发展起步较晚，1992 年后才陆续引入北美寿险精算师考试、英国精算师考试、日本精算师考试、北美非寿险精算师考试等，2000 年后，中国精算师考试体系逐步建立起来。目前，中国精算师考试的考点已增加到 15 个。2006 年 12 月，民政部批准中国精算师协会正式筹备成立。中国精算师协会的成立，必将进一步推动中国精算教育和精算师职业的发展，也迫切要求对当前的精算教育体系和精算师考试体系进行必要的改革，以尽快向国际精算师协会发布的精算教育标准看齐。

中国人民大学统计学院是国内较早开展风险管理与精算教育的大学之一。1992年统计学院就开始招收风险管理与精算专业方向的硕士研究生,1993年开始招收该方向的本科生,1996年招收了该专业方向的第一批博士研究生。2004年,经教育部批准备案,统计学院设立了独立的风险管理与精算学硕士学位点和博士学位点,标志着在风险管理与精算人才培养上,形成了学士、硕士、博士多层次、专业化的人才培养教育体系。其专业课程设置完全与国际接轨,涵盖了北美、英国和中国精算师初级课程考试的基本内容,教学大纲紧跟国际精算师协会公布的精算教育指南,同时根据学科发展的国际趋势,每年重新修订课程和教学大纲。在研究方面,设立了中国人民大学风险管理与精算中心。多年来,在寿险风险管理与精算、非寿险特别是汽车保险风险管理与精算、养老金、社会保障等领域取得了很多有影响的成果,进一步促进了风险管理与精算教育的发展。为适应我国精算教育改革与发展的需要,并体现与国际精算师协会的精算教育标准接轨,中国人民大学风险管理与精算中心精心组织编写了一套精算学系列教材,分两个阶段完成。第一阶段涵盖精算师考试初级课程的全部专业课内容,包括《金融数学》、《风险理论》、《寿险精算学》、《非寿险精算学》、《精算中常用的统计模型》5本教材,其中每本教材包含大量的练习题和解答。第二阶段涵盖精算师考试高级课程的全部内容,分寿险、非寿险、养老金、健康保险、社会保障、投资等不同系列。这套教材一方面可以满足各高校精算专业的教学需求,另一方面也可以作为参加各类精算师资格考试学员的学习参考资料,同时可以作为对精算学科有兴趣的同仁了解和学习精算的参考书。

这套教材的特点,一是在内容上涵盖了北美寿险、北美非寿险、英国、中国精算师考试最新的内容,同时紧跟国际精算师协会提出的精算教育标准,涵盖了国际精算教育大纲的基本内容;二是为了便于读者自学和教师讲授,我们在每本教材中配备了习题及详细解答;三是在写法上,力求把精算学的数理理论与实务结合起来,注意精算数学背后的实践意义,努力从实际意义上解释各种数学关系。

本套教材凝结了中国人民大学风险管理与精算中心全体教师的心血,特别是王晓军、孟生旺、黄向阳、王燕、肖争艳、肖宇谷等老师,他们为本套教材的编写付出了极大的艰辛,统计学院部分硕士研究生和本科生对教材中的习题解答和答案进行了验证,感谢他们为本套教材作出的贡献,同时也感谢中国人民大学出版社的编辑为本书出版付出的辛勤劳动。



前 言

精算应用中包含各种统计模型，这些模型独立出来都可以作为一门课程，但是对于基础的考试内容和简单应用，我们可以对它们进行适当的精简。经过对中国精算师考试大纲（2007）和 SOA 考试大纲（2007）的详细分析，我们把那些不适合归入传统精算课程的内容综合起来，以考试用书为基本目标，在包含所有有关考试内容的前提下，尽可能清晰地展现给读者。

本书由四个部分组成：（1）应用回归分析（1~3 章）；（2）时间序列分析（4~12 章）；（3）随机模拟（13 章）；（4）生存模型及其应用（14~18 章）。其中第一部分和第二部分覆盖中国精算师考试第 2 门课程（02 数学基础 II）的相应内容，在 SOA 体系中这两部分属于 VEE 的内容。第三部分覆盖 SOA 考试 COURSE 4 中关于随机模拟的内容，且第 13 章 1~5 节包含中国精算师考试第 5 门课程（05 风险理论）的有关内容。第四部分包含中国精算师考试第 6 门课程（06 生命表基础）的内容。本书由黄向阳编写第一部分，金阳编写第二部分，肖宇谷编写第三部分，宁威编写第四部分，肖宇谷负责整体审核。

本书四个部分基本上是彼此独立的，读者可以有针对性地选择阅读。

为了方便读者学习，我们将本书习题的全部答案放在人大经管在线（www.rdjg.com.cn）上，同时还提供了第 13 章随机模拟的例题选编，读者可以下载学习。

由于作者水平有限，书中不足之处敬请批评指正。

编著者

目 录

第一部分 应用回归分析

第 1 章 相关与一元回归分析	3
1.1 拟合与最小二乘法	4
1.2 两个变量的相关关系	8
1.3 一元线性回归	13
1.4 回归方程的应用	20
习题	25
第 2 章 多元线性回归模型	27
2.1 多元线性回归模型及其参数估计	27
2.2 多元线性回归模型的假设检验与区间估计	36
2.3 使用多元线性回归的注意事项	44
习题	52
第 3 章 异方差与序列相关的处理	56
3.1 异方差的存在与检验	56
3.2 广义线性模型和处理异方差的方法	63
3.3 自相关线性模型	73
习题	80

第二部分 时间序列分析

第 4 章 简单模型	85
4.1 简单外推模型	85
4.2 平滑技术	86
4.3 季节调整	88
习题	89
第 5 章 时间序列的概率结构	90
5.1 分布与特征统计量	90
5.2 平稳时间序列的定义	92
5.3 平稳时间序列的统计特性	93
5.4 白噪声序列	96
5.5 白噪声检验	97
习题	99
第 6 章 平稳时间序列	100
6.1 预备知识	100
6.2 AR 模型	103
6.3 MA 模型	116
6.4 ARMA 模型	120
习题	123
第 7 章 平稳序列建模	124
7.1 建模基本步骤	124
7.2 模型识别	125
7.3 参数估计	127
7.4 模型检验	132
7.5 模型优化	134
习题	136
第 8 章 时间序列的预报	137
8.1 线性预报	137
8.2 线性最小方差预报	140
8.3 修正预报	142

习题	144
第 9 章 非平稳时间序列	145
9.1 差分	145
9.2 ARIMA 模型	147
9.3 疏系数模型	152
习题	153
第 10 章 残差自回归模型	154
10.1 模型结构	154
10.2 残差自相关检验	155
10.3 模型拟合	158
习题	158
第 11 章 条件异方差模型	159
11.1 异方差的定义	159
11.2 异方差的初步识别	160
11.3 方差齐性变换	161
11.4 条件异方差模型	162
11.5 AR-GARCH 模型	165
11.6 GARCH 模型拟合	166
习题	169
第 12 章 多元时间序列	170
12.1 多元序列建模	170
12.2 单位根检验	171
12.3 协整	180
习题	182

第三部分 随机模拟

第 13 章 随机模拟	185
13.1 均匀分布随机数与伪随机	185
13.2 用反变换法产生一般分布的随机数	186
13.3 Cholesky 分解和多元正态分布的模拟	191
13.4 模拟样本的容量问题	192

13.5	模拟在精算模型中的应用举例	194
13.6	模拟在统计检验中的应用	198
13.7	用自助法计算估计量的均方误差	203
13.8	股票价格的对数正态模型和模拟	211
13.9	风险度量 VaR 和 CTE 的模拟	225
	习题	230

第四部分 生存模型及其应用

第 14 章	生存模型相关性质和生命表	235
14.1	生存模型	235
14.2	T 的分布函数	237
14.3	参数生存模型举例	240
14.4	条件概率的数字特征和截尾分布	243
14.5	随机变量的变换	247
14.6	生命表	250
14.7	尾龄的分布假设	255
14.8	选择生命表	257
	习题	258
第 15 章	完整样本数据和不完整样本数据情况下生存模型的估计	260
15.1	死亡时间	260
15.2	死亡的精确时间	261
15.3	死亡时间分组法	263
15.4	非完整样本数据情况下生存模型的矩估计	267
15.5	非完整样本数据情况下表格生存模型的极大似然估计	279
15.6	乘积估计量	285
	习题	288
第 16 章	参数生存模型的估计	289
16.1	完整数据的单变量模型	289
16.2	不完整数据下的单变量模型	291
16.3	参数模型的假设检验	293
16.4	参数模型中的伴随变量	294

16.5 大样本数据下年龄的处理及暴露数的计算	298
习题	303
第 17 章 人口统计	304
17.1 死亡和生育测度	305
17.2 静态人口模型	306
17.3 稳定人口模型	312
17.4 人口规划及人口普查应用	315
17.5 人口规划	317
17.6 人口统计的应用	320
习题	323
第 18 章 表格数据修匀	324
18.1 修匀的定义	324
18.2 移动加权平均修匀法	325
18.3 Whittaker 修匀	328
18.4 Bayse 修匀	331
18.5 二维 Whittaker 修匀	332
18.6 参数修匀	334
习题	343
附录：部分习题答案	344
参考文献	351



21世纪保险精算系列教材



第一部分

应用回归分析

第1章

相关与一元回归分析

本章重点

1. 掌握最小二乘法的计算公式以及在拟合中的应用
2. 了解相关关系和因果关系的异同
3. 掌握一元线性回归模型的假设和参数估计
4. 能够对需要变换的数据完成回归分析
5. 掌握虚拟变量的定义和解释

研究两个变量的定量关系并建立拟合这种关系的曲线是自然科学和社会科学中的常见手段。早期发展的是拟合方法，后来发展出能够检验误差影响的回归理论和算法，才算形成了完整的回归分析体系。本章按照方法的发展历史，从拟合和最小二乘法入手，介绍相关概念，然后介绍一元线性回归模型以及模型的参数估计和相关的假设检验，最后结合可以线性化的回归模型和虚拟变量介绍线性回归模型的应用。

注：本章使用的回归模型的符号并不统一（在正文部分的模型是 $y = \beta_0 + \beta_1 x + \epsilon$ ，在习题中经常出现的还有 $y = a + \beta x + \epsilon$ ，或者 $Y = a + \beta X$ ），主要是考虑到在 SOA 考题中出现的一般是后一种模型。读者只要记住区分回归直线的常数项（截距）和斜率，就不会发生混淆。

1.1 拟合与最小二乘法

考察变量之间的数量关系是现代科学方法的基本手段。它起源于伽利略，由牛顿加以完善，随后在物理学和工程实践中得到了广泛应用并取得巨大成功。从18世纪开始，这种方法逐渐影响到几乎所有学科。定量研究方法的基本程序包括：

- (1) 确认需要研究的变量。
- (2) 获得变量的一系列观测值。
- (3) 研究这些观测值之间的关系。
- (4) 建立变量之间的关系模型（一般表现为几个变量之间的函数关系）。

1.1.1 拟合的基本步骤

下面以工厂中产量对总成本的影响研究为例说明这种方法的基本步骤。

【例 1—1】 产量和成本的关系研究。为了研究产量和成本之间的关系，首先我们要收集一系列数据，而且必须是在一系列产量水平下的成本。这么做的理由是，按照普遍接受的理论，产量的增长会带来成本的增加。由此得到一系列产量和成本的观测数据的数对，如表 1—1 所示。

表 1—1 某种产品的产量和成本之间的关系

序号	1	2	3	4	5	6	7	8	9	10
产量	1.05	4.21	5.1	6.2	8.75	2.38	1.49	4.71	7.21	4.86
成本	3.52	15.73	15.07	24.81	32.58	11.32	7.52	17.23	24.33	20.96

在处理数据之前，用更直观的方式来展示这些数据往往可以帮助我们获得更清晰的整体印象。在两个变量的关系研究中，散点图是常用的工具，因为它可以把两个变量的数量关系表现在二维平面坐标系上。绘制两个变量的散点图用 Excel 就可以了，见图 1—1。

凭肉眼就可以判断出这 10 个点分布在一条直线的两侧，图 1—1 上的方程 $y=3.442x+1.4876$ 是用 Excel 的“添加趋势线”功能得到的。之所以用添加趋势线而不是回归分析模块来得到拟合直线，是因为在提出基于概率统计理论的回归模型之前，科学家和工程师已经习惯于对散点图拟合合适的曲线方程。所使用的算法是最小二乘法，最小二乘法和数理统计理论结合之后才产生出回归分析。一般认为这项工作主要归功于高斯和马尔可夫。

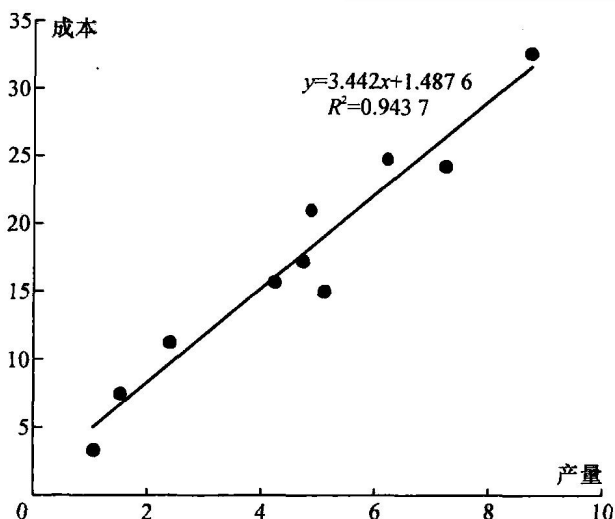


图 1—1 产量和成本散点图

从计算过程来看,给定描述两个变量之间数量关系的函数的形式,可以在不了解误差分布的条件下用最小二乘法估计该函数的各个参数。拟合与回归的发展历史可以印证这一点。回归分析的理论成熟于 19 世纪,但是科学家和工程师利用拟合已经得到了很多有用的经验法则,比如欧姆定律和波义耳的气体定律等。而勒让德在 1805 年提出最小二乘法的时候,也没有讨论误差的性质,更没有建立有关的模型和系数的显著性检验。这些问题到 1809 年才由高斯解决。回归分析理论的意义不是解决拟合函数的计算问题,而是指出:在存在各种误差的情况下,拟合得到的函数形式在多大程度上是可以信赖的。传统的拟合方法则没有给出令人信服的判断方法。因此我们先来看最小二乘法的理论和实现,然后讨论回归分析理论的假设检验问题,最后讨论回归分析模型的应用。

确定两个变量 (X, Y) , 考察这两个变量的一系列取值形成的数对,然后在平面坐标系内画出这些数对的散点图,进一步还可以为这些数对拟合出一条具有很强表达能力的趋势线,甚至作回归分析。这种研究思路在 21 世纪看来都是再普通不过的,在缺乏计算机和计量手段的 18 或 19 世纪却是很先进的方法。按照回归方法的发展过程,我们先看不考虑误差项性质的最小二乘法。

1.1.2 拟合中的最小二乘法

对两个变量 (X, Y) , 其中 X 不是随机变量。假定通过实验或者观察获得了 n 组观测数据,记为数对 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。如果根据理

论或者经验认为在 X 和 Y 之间存在着一定的线性关系，考虑到观测数据必然包含误差，就可以得到一系列包含误差的关系式：

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i=1, 2, \dots, n \quad (1-1-1)$$

式中， β_0 和 β_1 为待定的系数， ϵ_i 为未知的随机变量，在回归模型中会假定 $\epsilon_i \sim N(0, \sigma^2)$ ，且 $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ 相互独立，现在我们仅仅假定 ϵ_i 的期望为 0。在 $E(\epsilon_i) = 0$ 的假设下，有 $E(y_i) = \beta_0 + \beta_1 x_i$ 。写成函数之间的关系，就是 $E(Y) = \beta_0 + \beta_1 X$ 。每个 X 的取值 x_i 都有一个与之对应的随机变量 Y 的估计值 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ ，但是实际观测到的值是 y_i ，显然在 y_i 和 \hat{y}_i 之间存在着因误差导致的差异。

由于存在误差的干扰，式 (1-1-1) 不是普通意义上的线性方程组。实际上，这组关系式中只有两个待定的系数，而观测值的数量可能达到几十个，也就是说所有这些数据形成的方程组的基本特点是方程的个数多于未知数。通常采用普通最小二乘法 (ordinary least square, OLS) 对参数 β_0 和 β_1 进行估计。其准则是，对每一个样本观测值 (x_i, y_i) ，观测值 y_i 与其期望值 \hat{y}_i 之间的离差平方和最小。这样，估计 β_0 和 β_1 的问题就变成了一个最优化问题，即求 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ ，使得偏差平方和 $Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ 取最小值。

注意到求和项的平方运算，这就是最小二乘法得名的根据：所谓二乘，就是平方。在有了观测值以后，我们要解决的数学问题是：

$$\begin{aligned} Q(\hat{\beta}_0, \hat{\beta}_1) &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \\ &= \min \end{aligned} \quad (1-1-2)$$

$Q(\beta_0, \beta_1)$ 是关于 β_0 和 β_1 的二次函数，对 β_0 和 β_1 的偏导存在，通过令 Q 对 β_0 和 β_1 的偏导为零，可以得到 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 。

$$\begin{cases} \frac{\partial Q}{\partial \beta_0} \Big|_{\beta_0 = \hat{\beta}_0} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \\ \frac{\partial Q}{\partial \beta_1} \Big|_{\beta_1 = \hat{\beta}_1} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \end{cases}$$

经整理，得到关于 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 的二元一次方程组