

## 第四届“互动出版杯”数学中国

### 数学建模网络挑战赛

#### 承 诺 书

我们仔细阅读了第四届“互动出版杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：

参赛队员（签名）：

队员 1：

队员 2：

队员 3：

参赛队教练员（签名）：

参赛队伍组别：本科组

## 第四届“互动出版杯”数学中国

### 数学建模网络挑战赛

#### 编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

1037

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

# 2011 年第四届“互动出版杯”数学中国 数学建模网络挑战赛

题 目 续保因素分析及电销业务占有率预测

关 键 词 方差分析 权重 logistic 续保率 电销占有率

摘 要:

本文基于方差分析的原理对影响续保率的因素进行分析，给出了对续保率的影响强弱判定模型；建立了电销发展的 logistic 模型以分析电销的引入对于保险企业的影响，预测电销业务对传统销售方式的取代程度。

首先，我们用方差分析法，对目标客户数，承保车辆的使用性质，新车的购买价格和承包车辆的销售渠道，承保车辆的年龄，承保车辆的出项次数以及承保车辆的牌加权，建立权重模型，定量的描绘了他们对于续保率的影响强弱。

其次，在分析电销业务对企业的影响时，我们从企业的整体的销售状况和电销所占的比例进行分析，分别研究电销引入前后保险企业各结构参数，进行对比，其额定除主要因素和次要因素，从而阐述电销对保险企业的影响。在对电销业务取代传统方式的预测上，我们类比人口的 logistic 模型对电销业务在每年的各个季度销售中所占的比重进行拟合，定量的描述了电销业务所占的份额和发展趋势，最终，电销业务所占的比重将趋于稳定，同时传统业务和其他新兴业务也趋于稳定，由此得出电销对传统方式的取代程度。

参赛队号 #1037

所选题目 c 题

参赛密码  
(由组委会填写)

## 一、问题重述

近些年来，国内汽车市场的销售火爆，于此同时汽车保险的数量也逐年增加，尤其是在交强险实施之后，汽车保险在保险业中所占的比重越来越大。

保险公司在保障自己的市场占有率的时候，留住老客户是很重要的一部分的内容，也就是续保率。影响续保率的因素有很多，其中有新车购买价格，承保渠道，承保车辆使用性质，承保车辆品牌，承保车辆年龄，承保车辆出险次数等。

问题 1：建立合理的数学模型，说明影响续保率的因素。

电销业务是依托近年来信息产业迅速发展而兴起的一种新型业务，它因其保费较低而让大多数的人感受到了实惠，所以在一开始，电销就表现出了强劲的势头。我们现在知道保险公司中电销的数据。

问题 2：建立数学模型，全面评估电销业务的推广对保险企业的影响，预测电销的方式将在多大的程度上取代传统的方式。

## 二、问题分析

### 1.1 续保率的定义

续保率是一个保险公司保证其业务量的重要的一环，在市场竞争激烈的今天显的尤为重要。

续保率 = (已续保的用户 / 当年到期的车辆数) \* 100%

其中当年到期的车辆数 = 当年到期的目标客户车辆数 + 当年到期的非目标客户车辆数  
目标客户是：题目所给数据中的目标客户是指出险次数 0-1 次且赔付金额占保费收入 60% 以下的客户。

### 1.2 影响续保率的主要因素

在已给出的影响续保率的主要因素中有：承保车辆出险次数的统计，承保车辆的年龄，承保车辆的品种，承保车辆的使用性质，车辆的承保渠道，新车的购买价格。

承保车辆的出险次数：没有出过险的客户的续保意愿明显会较出过险的客户的续保意愿低上一些，有没有出过险直接影响了客户的续保意愿，而那些从没出过险的客户正是目标客户是保险公司要争取的。

承保车辆的年龄：随着车辆使用年限的增大，出现事故的概率也必然会增多，所以车辆使用年限较多的客户续保的概率应该会更大些，是续保因素中不可或缺的。

承保车辆的品种：不同牌子的车辆，其质量的好坏，价值的高低，售后服务的好坏都会影响客户的续保率，起对续保率的影响较为复杂，不同品种的续保率可能会有很大差别。

承保车辆的使用性质：在使用性质里面大致可以分为商用车和非商用车。商用车主要包括城市公交，公路客运，营业货车等；非商用车主要有非营运货车，家庭自用，党政机关客车等。对于商用车，由于其行驶里程较多，所以其出现事故的概率也较大，所以续保的意愿应较非商用车高。所以使用性质应该是影响因素里比较重要的一个内容，在续保率中起着重要的作用。

车辆的承保渠道：其中主要有电话销售，交叉销售和车商渠道。不同的销售渠道对续保率也会有着影响。

新车的购买价格：这个因素对续保率的影响时直接的，大多数人都会选择给较贵的车如保，来减少在发生意外之后自己的损失。

### 1.3 数据有效性的判断

对于车辆品牌这个因素，我们对其进行了剔除的处理。主要原因是：车辆的品牌过多，并不有利于进行统计；对于个别品牌的车辆来说，销量过小以至于无法准确的显示出续保率；对于不同品牌的车辆的质量问题，我们认为可以归结到车辆的出险次数中去；对于好牌子与差牌子的车，我们认为可以归结到新车购买价格里去，综上所述，我们对车辆的这个因素做了剔除的处理。

对于验证数据的有效性，由于保额应该为非负数，即使有退款的也应该不会有负的出现，故0万元以下的数据作为坏值。

#### 1.4 续保率的分析

在以上这些影响续保率的因素中，我们选取承保车辆的出现次数，承保车辆的年龄，车辆的承保渠道，承保车辆的使用性质和新车的购买价格来进行分析。我们打算使用方差分析法算出每个因素的权重，权重的大小表征了该因素对于续保率的影响程度，然后来进行一个比较，从而确定那个因素的影响较大。

对于车辆品牌这个因素，我们对其进行了剔除的处理。主要原因是：车辆的品牌过多，并不有利于进行统计；对于个别品牌的车辆来说，销量过小以至于无法准确的显示出续保率；对于不同品牌的车辆的质量问题，我们认为可以归结到车辆的出险次数中去；对于好牌子与差牌子的车，我们认为可以归结到新车购买价格里去，综上所述，我们对车辆的这个因素做了剔除的处理。

#### 2.1 电销业务增长的分析

电销业务是新兴的一种销售方式，其的增长势必会对原有的销售方式进行压缩，我们需要分析的就是电销对传统的销售方式进行了何种程度的压缩。

电销在起初出现的一段时间势必会经过一个增长期，然后会趋于平稳，各种销售方式都会趋于稳定。这种变化方式与人口增长模型很类似，我们可以引入logistic模型，通过已给数据讨论其参数已达到对电销的预测。

在这个过程中，我们还要注意到，出了电销这种新兴的销售方式以外还有其他一些的新型方式的出险，且占据了大量的市场份额。在分析电销对传统的销售方式的影响时要注意处理其他新兴的销售方式对传统的销售方式的压缩。

### 三、基本假设和参数说明

#### 1. 基本假设

假设各因素之间是相互独立的；

假设续保不受除锁讨论因素的其他因素影响；

假设电销是从2010年第2季度出现并发展的，且当时的值电销用户数为所有参保的5%，电销推销的客户是随机的。

#### 2. 参数说明

$X_i$ 、 $Y_i$ 两种不同因素中各成分的续保率，其平均值为。

$C_i$ 各个因素的方差。

$M$ 为电销对于传统销售方式的期待成程度

$k$ 值为电销的增长率

$t$ 是相对于初始季度的季度数，2010年第二季度 $t=0$

$R$ 为相关系数

## 四、建立模型

### 1.1 方差分析法

方差分析法是一种简便、灵活而实用的决策方法。它是把复杂问题中的各因素划分成相关联的使之条理化的多目标、多准则的决策方法，是一种定量分析与定性分析相结合的有效方法。

在横向比较的过程中，利用方差分析法可以对不同因素在续保的权重由一个清晰的了解。首先通过判定目标续保率和总续保率之间的线性关系，来确定出目标续保率对总续保率的影响。

### 1.2 权重理论具体原理及实验步骤

首先把各个因素分离出来，算出各个因素的参量的利差平方和作为影响续保率的权重。然后进行横向的比较，判断出哪个对续保率的影响较大。

计算公式为：

由此算出各个因素的权重 $C_i$ 。

2.1 我们首先从给出的数据中确定出电销增长率，类比人口增长模型，用，迭代发解方程，估算出M的值。

### 2.2 电销业务对保险公司的影响

首先我们可以对比两个季度的数据，通过电销引入前后数据的变化来反映对保险企业的影响。

然后，可以直观的根据已经给出的数据来绘制出表格和图形进行比较，得出在2010年第一季度和2011年第一季度的电销的主要增长的方向。在交强险和商业险分别绘制2个图形，比较出电销业务主要是在那一块增长的，在哪里增长的更大一些。绘制出电销业务里各个的构成，结合上面确立的电销增长模型来分析电销的业务范围的变化。

如上图是电销业务中，新保和续保占得比例。

## 五、计算模型

问题一、

1、对于因素“承保车辆年龄”，我们首先分月份绘出到期总车数与车辆年龄的折线图如

下：

由上图可以看出，各个月份到期总车数随车龄的走势基本一致

我们又画出了到期目标总车数与车龄的折线图如下：

由上图可以看出，每个月份的出险车数-出现次数的走势基本一致

由上，我们可以把各个月份的平均值拿出来研究，根据按照承保车辆年龄统计中的当年到期车辆续保率和目标客户续保率我们计算出了当年到期续保车数和目标到期车辆数，并且计算出了总的续保车辆和总的目标续保车辆，得到如下表格：

表格 1 总续保情况

车龄	2011 .3	2011 .2	2011 .1	2010, 1 2	2010. 1 1	2010. 1 0	2010 .9	总 计	续 保 率
0-1 年	3497	1966	1186	11722	10523	9948	9161	48003	0.265
1-2 年	2048	1176	698	5638	5289	5062	4451	24362	0.304
2-3 年	1182	580	316	2958	2786	2651	2408	12881	0.237
3-4 年	759	402	226	2127	2061	1976	1783	9334	0.230
4-5 年	468	238	162	1348	1268	1209	1142	5835	0.197
5-6 年	429	214	117	1010	957	907	824	4458	0.207
6-7 年	282	140	102	582	549	513	472	2640	0.214
7-8 年	139	64	49	352	339	323	299	1565	0.205
8-9 年	58	28	21	215	205	195	176	898	0.181
9-10 年	35	21	15	122	111	107	98	509	0.172
10 年以 上	79	40	20	248	231	220	207	1045	0.252

表格 2 目标客户续保情况

车龄	2011 .3	2011 .2	2011 .1	2010, 12	2010. 11	2010. 10	2010. 9	总计	续 保 率
0-1 年	2416	1336	820	7268	6538	6202	5659	30239	0.235
1-2 年	1676	949	567	4490	4159	3896	3369	19106	0.295
2-3 年	1025	491	262	243	2289	2146	1941	10593	0.230
3-4 年	658	350	196	1795	1749	1646	1474	7868	0.224
4-5 年	416	210	143	1149	1073	1020	964	4975	0.189
5-6 年	374	186	99	877	826	780	709	3851	0.190
6-7 年	253	127	90	506	481	443	407	2307	0.199
7-8 年	121	56	44	310	299	283	265	1378	0.205
8-9 年	55	27	20	195	185	178	162	822	0.202
9-10 年	35	21	15	109	98	93	86	457	0.179
10 年以 上	73	36	18	232	215	202	190	966	0.173

根据上表，绘制总续保率和目标客户续保率的折线图如下：



上图表明了总续保率和目标续保率的关系：两者差别很小。

假设两者之间线性关系，即  $X=Y$ 。

下面通过计算它们的相关系数来验证他们相关的显著性，

参数定义如下：

参量	0-1 年	1-2 年	2-3 年	3-4 年	4-5 年	5-6 年	6-7 年	7-8 年	8-9 年	9-1 0年	10年 以上
总续保率	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
目标续保率	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11

相关系数计算公式为：

用 MATLAB 编程（见附录 1）计算得： $0.7636 > R_{0.05(9)} = 0.602$

所以若以 0.05 作为显著性水平，我们有 95% 的把握说明  $X_i = Y_i$ ，即它们之间有很强的线性关系。

这里的说明目标客户对于总的续保率的贡献不大，我们对于其他几个影响因素也用同样的方法处理，得到如下数据

按车辆年龄统计： $R = 0.8763 > R_{0.01(9)} = 0.735$

承包车辆使用性质： $R = 0.9964 > R_{0.01(7)} = 0.798$

按新车购买价格： $R = 0.9984 > R_{0.01(5)} = 0.847$

按承包渠道： $R = 0.9402 < R_{0.05(2)} = 0.950$

同样的结论：目标客户对于总的续保率影响不大，在以下讨论个因素时均以总的续保率数据为依据。

2、作出每一种因素的几个月来的平均续保率折线图如下

在此因素中，我们可以看到：使用年龄续在 0-1 年和 2-3 年的续保率差别很大，而出使用年龄在 4 年以上的续保率无明显差别，基本和使用年龄无关。

在此因素中，我们可以看到：出险 0 次和 1 次的续保率差别很大，而出险两次以上的续保率无明显差别，基本和出险次数无关。

此图表明：城市公交的续保率明显高于其他，主要原因是城市公交行车时间最长，只要

在上班时间就一直在运行；而行车时间最短的非营业货车最低，。。

在此因素中，我们可以看到：购买价格在 20 万以下的续保率差别比较大，而购买价格在 20 万以上的续保率无明显差别，基本和购买价格无关。  
我们还利用公式 编程（程序见附件 2）计算出了以上各个因素的利差平方和如下：

因素	车龄 C1	出险次数 C2	车辆使用性质 C3	承购渠道 C4	购买价格 C5
离差平方和	0.0955	0.0140	0.2322	0.0314	0.024

对比各幅图我们还能直观地得出结论：

- （1）、权位最大的为按照承保车辆的使用性质，即，该因素对于续保率的影响最大，权重最小的是按照承保车辆购买价格，即该因素对于续保率的影响最小；
- （2）、按照承保车辆年龄和承保渠道统计的目标客户权和总权差别较大，分别为 50% 和？，这说明承保车辆的年龄对于新老客户的续保影响很大，但是承保车辆的权明显比其他因素的权位小很多，其并不是主要因素

## 问题二

- 1、我们首先对两个季度的数据分别分析，绘制出它们的饼状图如下：

而横向比较图如下：

读上面的饼状图，我们发现，10年第一季度和11年第一季度的保额比例、保单性质比例、险种类别比例所差最大都不超过16%，我们就说它们对于电销的引入变化迟钝，即电销的引入没有对上述三类结果构成产生较大的影响。

1、接下来我们研究电销的引入对于展业方式的影响  
2011年第一季度：

2010年第一季度：

对比可知：

(1)、电销的引入对传统销售方式有了很大的压缩，交强险和商业车险中营销员的压缩比为：31.5%和56.7%，专业代理的压缩比例分别为：76.1%和63.9%；

(2)、不仅传统的被压缩，在引入电销后还产生了一些新的销售方式，像汽车销售行、运输行业等。

(3)、2011年第一季度电销数550个，根据假设我们算得增长率 $k=8.6\%$ 。

(4)、随着电销的出现，更多的销售方式（像汽车销售行、运输行业等）也逐渐出现，并且还占有了较多的百分比，比电销更强势。

### 3、对电销未来走势的预测

对比人口增长模型，我们构造出电销增长logistic模型如下：

由2中数据 $k=8.6\%$ 我们得到方程：

由运用牛顿迭代法（程序见附件3）求得的近似解为： $M=550$  此值刚好等于2011年的电销量。

由此我们得出电销在所有销售方式中能占的最大比例为6.4%，即在6.4%的程度上取代传统销售方式。尽管电销比传统方式方便，但我们通过图像也发现，伴随着电销的出现，其他一些新兴的销售方式（像汽车售行、运输行业等）也随着出现，它们比电销更方便、更强势。

## 六、企业决策

- 1、问题二中我们已经得到结论，现今电销已达到饱和，说以以后对电销小没必要再投入太多的精力，而应该把重心转移至怎样开拓新的更方便的销售方式上，特别是现在互联网的更加普及。

## 七、模型的评价及优化

### 1、模型的优点：

（1）模型的合理性，模型的建立是在对样本数据进行充分挖掘的基础之上的，通过数据之间的内在关系观察计算，绘出图表，提炼出各个因素对于续保率的影响。

（2）模型简单、直观，问题一的模型通过计算各个影响续保率的因素的权位，说明了各因素对于续保率的关联程度；模型二通过logistic模型较准确地预测了未来电销取代传统销售方式的程度。

（3）模型利用excel作了大量的数据处理，模型的可靠度高。

### 2、模型的缺点：

（1）给定的数据中的某些负值，可能有些实际意义，但我们帮它作为了坏值处理。

（2）模型二虽然预测较准确，但在建立模型时更多的是依据想象，没有太多的依据。

### 3、模型的改进：

（1）从按车辆年龄统计、承包车辆使用性质、按新车购买价格、按承包渠道、车辆出险次数与续保率的图像上我们还可以发现一个横有趣的现象，凡是和数字有关的因素（像车辆年龄、新车购买价格、出险次数）的图像走势都有一个共同的特点，那就是，当数字很小时波动幅度很大，随着数字的增加而又趋于平缓，我们猜想这种现象反映了大众的心理，这是一个值得探讨的问题；

（2）在研究电销对保险企业的影响时，如果能够引入对营业员工资状况的数据进行分析，会更好、更直观地从盈利方面进行表述，苦于缺乏数据和平均赔付率的难预测性，我们就没做；

（3）logistic模型只是一个非常近似模拟，我们本打算用个3阶转移矩阵，来描述电销、传统模式还有新兴模式的互相作用关系，不过得到很复杂的方程组，难以很容易求得，不过那却是一个很不错的方法。

### 八、参考资料：

- [1] 徐全智 吕恕，《概率论与数理统计》，高等教育出版社，2004年第一版
- [2] 徐全智 杨晋浩，《数学建模》，高等教育出版社，2008第二版
- [3] 姜启源 谢金星，《数学建模案例选集》，高等教育出版社，2006年第一版
- [4] 周品 赵新芬，《MATLAB数学建模与仿真》，国防工业出版社，2009年第一版

### 九、附录

附录1，求解相关系数的程序：

```
function xiangguanxishu
x1=input('请输入行向量X1:');
y1=input('请输入行向量Y1:');
m=size(x1);
n=m(2);
sumx1=sum(x1);sumy1=sum(y1);
averx1=sumx1/n;avery1=sumy1/n;
sumxy=0;sumx2=0;sumx22=0;sumy22=0;
for i=1:1:n
    sumxy=x1(i)*y1(i)+sumxy;
    sumx2=x1(i)^2+sumx2;
    sumx22=(x1(i)-averx1)^2+sumx22;
    sumy22=(y1(i)-avery1)^2+sumy22;
end
rotx1=sqrt(sumx22);roty1=sqrt(sumy22);
bu=sumxy-n*averx1*avery1;
bl=sumx2-n*averx1^2;
rot=bu/(rotx1*roty1)
```

附录2，计算离差和的程序：

```
function xiangguanxishu
m=size(x);
y=0;
n=m(2);
x1=sum(x)/n;
for i=1:1:n
    y=(x(i)-x1)^2+y;
end
x=input('请输入行向量X:');
```

y

附录3，牛顿法非线性方程 $f(M)=0$ ， $x_0$ 为迭代初值， $fname$ 和 $dfname$ 分别表示 $f(M)$ 和 $f(M)$ 的导数：

```
function newton(fname, dfname, x0, e)
if dfname==''
    dfname=diff(fname)
end
if e==''
    e=1e-4;
end
x=x0;
x0=x+2*e;
while (x0-x>e)
    x0=x;
    x=x0-eval(fname)/eval(dfname);
end
```