

硕士研究生学位论文

|  |  |
| --- | --- |
| 题目： | **P2P租车平台商业车险费率** |
|  | **厘定方法与实证研究** |

|  |  |
| --- | --- |
| 姓 名： | 肖陆镝 |
| 学 号： | 1401210784 |
| 院 系： | 软件与微电子学院 |
| 专 业： | 计算机技术 |
| 研究方向： | 电子商务与物流 |
| 导师姓名： | 李杰教授 |

二〇一六年四月

版权声明

任何收存和保管本论文各种版本的单位和个人，未经本论文作者同意，不得将本论文转借他人，亦不得随意复制、抄录、拍照或以任何方式传播。否则，引起有碍作者著作权之问题，将可能承担法律责任。

# 摘要

随着社会资源约束趋紧、环境污染日趋严重，P2P租车公司这类以共享经济模式运营的企业迅猛地发展起来。由于P2P租车平台会在每辆车上安装OBD设备采集车辆行驶数据，具有天然的大数据资源，因此对这部分数据的挖掘利用能够成为推进我国车险费率市场化进程的一个突破口。

本文试图通过对文献的学习以及定性定量相结合等方法，找出一种能够区分被保险人出险大小的方法，为更加精准的车险定价提供理论依据，提升保险公司的盈利能力。

首先，通过对我国车险价格厘定现状的研究以及对P2P租车平台商业车险特点的分析设计出保费的厘定模式，即日保费=日租金\*（基础费率+不计免赔额费率），基础保费的大小根据从人静态因素风险大小来拟定，不计免赔额费率依据驾车行为评分的高低来拟定。

其次，从多个数据维度中筛选出地区、性别、年龄、驾龄四个指标来进行静态从人因素风险评价，以运用服从Tweedie分布的广义线性模型取得了很好的模型拟合效果。

再次，通过对动态样本的分析，本文建立了多维度驾驶行为分析指标，并从中筛选出夜间行驶时间、每百公里急减速次数、行驶总里程、最大连续驾驶时间四个指标来对被保险人驾车行为进行打分。运用熵权法对各指标进行赋权后，建立了驾车行为评分表，并经过了有效性检验。

最后，通过各风险因子的分析，我们发现地区、性别、年龄、驾龄这四个因素对被保险人风险大小的衡量均有显著作用。地区方面，北京出险风险大于上海；性别方面，男性出险风险远大于女性；从年龄来看，21~24岁的被保险人风险最高，40岁以上的被保险人风险其次，而25~29岁年龄段的风险相对较低；从驾龄来看，驾龄小于1年的被保险人风险最大，其次是驾龄在4~7年的被保险人，而7年以上驾龄的被保险人，其风险会随着驾龄的增加而降低。驾车行为中，每百公里急减速次数这一行为因子对出险风险的影响最大，夜间行驶时间、行驶总里程以及最大连续驾驶时间这三个行为因子对风险的高低也有比较显著的影响。

关键词：P2P租车平台商业车险，静态从人因素，动态从人因素，广义线性模型，熵权法。

Auto insurance pricing and epirical reserch on P2P car rental company

Xiao Ludi (E-commerce)

Directed by Professor Li Jie

# ABSTRACT

With the social tight resource constraints, environmental pollution is becoming more and more serious, P2P car rental companies this kind to share the economic mode of operation of the enterprise rapid development. Because P2P car platform will in every car installed OBD equipment acquisition vehicle driving data, with natural data resources, so in this part of the data mining can be used to become to promote a breakthrough in our country insurance premium rate marketization process.

This article attempts through the literature study and qualitative and quantitative analysis method of combining, find a can distinguished by the size of the insurance of danger, provides the theoretical basis for the more accurate auto insurance pricing, improve the profitability of insurance company.

First of all, through the research of our country insurance price determined status and premium rating model analysis and design of the P2P platform rental commercial insurance characteristics, daily rent premium = rent fee\* (basic rate + excluding free compensation rates), the size of the base premium according to static factors risk size from to develop, regardless of the free compensation rates based on driving behavior score level to develop.

Secondly, from the multiple data dimensions screened area, gender, age, driving age four indicators to static from the human factor risk assessment to use obey the Tweedie distribution of generalized linear models and achieved good effect of model fitting.

Thirdly, through the analysis of dynamic samples, this paper establishes the multi dimension driving behavior analysis indicators, and the selected driving at night time, per 100 km rapid deceleration times, traveling mileage, the maximum continuous driving time four indicators to the insurer driving behavior points to play. Using the entropy of each index weight, a driving behavior scale, and through effective inspection.

Finally, through the analysis of the risk factors, we found area, gender, age, driving the four factors on the size of the insurance people risk measure has significant effect. Area, danger risk in Beijing than Shanghai; gender, danger risk in men far more than women; from the point of view of age, 21-24 years old is risk insurance were highest risk insurance were secondly, and 25 to 29 years of age the risk of relatively low is over 40 years old the adherends; from the point of view of the driving experience, driving for less than 1 year is insurance risk maximum, followed by driving experience in 4 ~ 7 years of the insured, and more than 7 years of driving experience is the insurer, the risk will decrease with the increase of driving experience. Driving behavior, per 100 km rapid deceleration times this behavior factor has the greatest effect on the risk of danger, the night of the travel time, driving mileage and the maximum continuous driving time these three behavioral factors of risk level are more significant influence.

KEY WORDS: Auto insurance of P2P car rental company, Variable static factors about drivers, Variable dynamic factors about drivers,Generalized Linear Model,Entropy weight.

# 目录

[摘要 I](#_Toc448295843)

[ABSTRACT II](#_Toc448295844)

[目录 IV](#_Toc448295845)

[第一章 绪论 6](#_Toc448295846)

[1.1 选题背景及意义 6](#_Toc448295847)

[1.2 研究目标、研究内容以及创新点 6](#_Toc448295848)

[1.3 论文组织结构 7](#_Toc448295849)

[第二章 车险定价理论与文献综述 8](#_Toc448295850)

[2.1 车险费率厘定概述 8](#_Toc448295851)

[2.1.1 车险费率厘定基本概念 8](#_Toc448295852)

[2.1.2 车险费率厘定原则 8](#_Toc448295853)

[2.1.3 车险费率厘定中关键指标—索赔频率与索赔次数 9](#_Toc448295854)

[2.2 车险费率厘定的基本方法分析 10](#_Toc448295855)

[2.2.1 总平均费率厘定法 10](#_Toc448295856)

[2.2.2 分类风险费率厘定法 10](#_Toc448295857)

[2.2.3 个体风险费率厘定法 12](#_Toc448295858)

[2.3车险费率厘定的风险因子选择 13](#_Toc448295859)

[2.3.1 从车静态费率模式 13](#_Toc448295860)

[2.3.2 从人静态费率模式 15](#_Toc448295861)

[2.3.3 从人动态费率模式 16](#_Toc448295862)

[第三章 P2P平台商业车险特征分析及厘定模型 22](#_Toc448295863)

[3.1 P2P租车平台商业车险特征分析 22](#_Toc448295864)

[3.1.1 我国当前商业车险费率厘定方法 22](#_Toc448295865)

[3.1.2 我国当前商业车险费率市场化改革进程 22](#_Toc448295866)

[3.1.3 P2P商业车险特征与厘定方法分析 23](#_Toc448295867)

[3.2 车联网的构成以及在保险中的作用 24](#_Toc448295868)

[3.2.1 车联网构成简述 24](#_Toc448295869)

[3.2.2 车联网设备在保险中的作用 25](#_Toc448295870)

[3.3 P2P商业车险厘定步骤分析 26](#_Toc448295871)

[第四章 从人静态因素的风险度量 28](#_Toc448295872)

[4.1 样本数据说明 28](#_Toc448295873)

[4.2 模型的选取 28](#_Toc448295874)

[4.3 指标的选取与数据的预处理 28](#_Toc448295875)

[4.4 损失分布的确定 33](#_Toc448295876)

[4.5 模型结果及分析 35](#_Toc448295877)

[第五章 从人动态因素的风险评价 37](#_Toc448295878)

[5.1 样本数据说明 37](#_Toc448295879)

[5.2 模型的选取 37](#_Toc448295880)

[5.3 驾驶行为指标的选取与数据的预处理 38](#_Toc448295881)

[5.4 驾驶行为指标的再筛选与指标权重的计算 42](#_Toc448295882)

[5.5 建立驾车行为评分模型，对驾车行为进行打分 43](#_Toc448295883)

[5.6 检验评分模型的有效性 44](#_Toc448295884)

[第六章 结论 48](#_Toc448295885)

[参考文献 49](#_Toc448295886)

[附录 51](#_Toc448295887)

[A 重要代码 51](#_Toc448295888)

[B 数据处理&分析工具 61](#_Toc448295889)

[致谢 62](#_Toc448295890)

[北京大学学位论文原创性声明和使用授权说明 63](#_Toc448295891)

# 第一章 绪论

## 1.1 选题背景及意义

对于解决社会资源约束、环境污染日趋严重的问题，共享经济功不可没。而其中最具代表性的就是出行行业。除了大家所熟知的滴滴这类拼座服务（司机是车主），宝驾这类主打P2P租车（司机是租客）公司的业务也迅猛地发展起来。

由于汽车购置金额较大、风险影响因子较多且事故发生频率较高，除了交通强制险之外，大部分私家车所有者还会购买商业车险。而对于P2P平台的车来说，除了车主必买的交强险之外，商业车险的购买方已从车辆所有方转移到了租客身上，此时保险中的信息不对称性以及逆向选择问题便会更加突出，传统商业车险定价体系已不再适用。另一方面，P2P租车公司出于安全性的考虑会为其平台上的每一辆车安装OBD设备，它可以提取到车辆的实时位置、速度、方向等信息。通过这个天然的数据入口，我们可以分析出租客的驾驶行为，以此来判断租客的驾驶风险。

本文首先根据数据指标性质不同将数据进行分类，分为从车、从人（静态、动态）因素两大类，通过定性定量相结合的方法设计出P2P商业车险厘定步骤，并进行实证检验，为P2P租车商业车险进行用户风险衡量、精准定价提供理论依据。与此同时，随着我国商业车险定价的逐步放开，本文对于车险行业的精准定价也有一定的借鉴意义。

## 1.2 研究目标、研究内容以及创新点

本文首先对我国普遍采用的商业车险厘定方法进行梳理分析，并在此基础上对比P2P租车平台商业车险目前所采用的计费标准以及根据P2P租车平台的自身特点找出车险厘定因子筛选原则与厘定步骤。其次，通过分析P2P平台可以获取到租客的准确信息，将这些数据指标分为从车、从人两大类。同时，根据租车平台给每辆车安装OBD设备可以获取到汽车驾驶人的动态数据，再将从人信息分为静态因子和动态因子两部分。

在确定风险因子后，设计出车险费率的厘定方法与步骤。由于交通事故出险事故中，仅有5%-10%的因素是与从车因素有关，而车辆购置价格又是决定出险金额高低很重要的一个决定因素。因此，将租车金额（与车辆购置价格高低高度相关）作为保费的一个乘数，而通过从人因素风险高低的判别来决定费率的大小。

在进行从人因素风险大小判别时，通过对现有车险定价方法的研究，分别为从人中的静态、动态驾驶行为因子选择广义线性模型、熵权法进行用户风险大小的拟合与预测。同时，根据P2P商业车险特点，将动态因子的风险因素放入免赔额费率大小的厘定中。

最后，本文试图建立一套相对完整的P2P租车平台商业车险厘定方法，能够更好地区分出租客风险大小并对其驾车行为起到一定的正向激励作用，从而给P2P商业车险的定价提供理论支持。与此同时，为我国车险费率市场化的开展提供实证研究的支撑。

本文的创新点主要体现在以下几个方面：第一，由于P2P租车模式在我国才刚开始起步，以往文献中没有专门针对P2P车险来进行研究的，本文将重点聚焦在P2P商业车险这个保险品种中的细分产品，填补了这部分内容研究上的空白；第二，本文在传统的车险费率厘定静态因子中加入了驾驶行为这一动态因子，为车险费率的厘定提供了新的思路。与此同时，也为我国大数据的实际运用提供了理论支持。

## 1.3 论文组织结构

全文总共分为六章，具体结构安排如下：

第一章 阐述了研究本文的背景及意义，P2P租车公司商业车险的自身特点，介绍了本文的研究目标、内容、创新点和组织结构。

第二章 相关理论的研究与介绍。首先介绍了车险费率厘定的基本概念，然后对费率厘定的基本方法进行了梳理，接着分别对从车费率模式、从人费率模式以及从人动态费率模式进行了研究与介绍，其中在从人动态费率模式这部分介绍中再次对驾驶行为理论做了较为详尽的梳理。

第三章P2P平台商业车险特征分析及厘定模型。首先简述了我国目前车险费率的厘定方法，之后对P2P平台商业车险这个市场细分产品进行了深入分析，总结了其独有的特征，并设计出针对P2P租车平台的商业车险的厘定方法。

第四章 从人静态因素的风险度量。首先对此次研究所采集到的静态从人因素样本加以介绍，接着从多个维度因子中选取加入模型作为自变量的风险因子，然后评价不同分布模型的拟合效果并确定损失分布，最后对模型所得数据的结论进行实际意义的阐释，总结具有不同特征的被保险人风险的大小。

第五章 从人动态因素的风险评价。首先对此次研究的动态从人因素样本数据进行简单介绍，然后基于原始样本数据进行了中间变量、最终指标变量的搭建与计算。接着对出险样本与未出险样本进行了指标变量的比对，并选出了加入评分模型中的动态行为指标。之后，运用熵权法对各个指标进行赋权，并设计出驾驶员驾车行为评分表。最后，在对评分结果与出险概率大小的比对中检验了评分模型的有效性。

第七章 结论。总结了本次论文所得到的结论及不足之处，并提出展望。

# 第二章 车险定价理论与文献综述

## 2.1 车险费率厘定概述

保险定价是指保险公司根据经营目标制定保险产品的价格。在实际中保险定价又包括两方面：车险费率的精算厘定和市场销售价格厘定。本文中的车险费率厘定是指车险费率的精算厘定。其中，精算费率厘定是根据公司的长期利润目标，基于保险原理，并在考虑已经承保的业务基础上进行厘定。

车险是财产保险的一种，它是以机动车辆本身及驾驶行为所涉及相关人员为标的，对于自然灾害或者意外事故所造成的财产损失或人身伤亡具有赔付责任的一种保险。根据保险标的的不同，车险又可以分为交强险、第三者责任险、车辆损失险三种主险，以及全车抢盗险、司机座位险、乘客座位责任险、玻璃单独破碎险、车身划痕损失险、涉水行驶险、不计免赔7种附加险。

保险人是承担保险标的受损风险的一方，一般是经营车险业务的保险公司，负责保险合同的拟定、保费的收取以及出险的赔偿。投保人是与保险人签订保险合同的一方，负有支付保费的义务。被保险人指依据合同约定其财产或者人身受保险合同保障的一方，享有保险事故发生后保险金的请求权。

### 2.1.1 车险费率厘定基本概念

保险费，是指投保人将标的风险转移给保险人所支付的费用，主要包括纯保费、费用附加、和利润附加。纯保费用于补偿保险公司在未来的期望赔款；费用附加用来补偿保险公司经营保险业务的各种必要的费用支出；而利润附加是经营保险业务所得到的收益。

保险金额，指保险合同中规定的保险人在承担赔偿保险金的最高限额，即保险标的的实质保险金额。

车险费率，指单位保险金额的保费。三者之间的关系可以表示为：

保险费=保险金额\*车险费率。

### 2.1.2 车险费率厘定原则

投保人将标的风险转移给保险人时所支付的保险费等于保险金额乘以车险费率，它反映了保险人所承担的风险大小。由于针对具体的保险标的保险金额是固定的，而车险费率的大小是衡量标的风险大小的关键指标。因而，车险定价的关键在于车险费率的厘定。

保险费率厘定的厘定首先要遵循以下四个原则来进行：

1. 保险偿付原则

保险偿付是指保险人按照保险费率向投保人收取的保险费，必须足以应付保险金给付及各种经营费用的需要。亚当.斯密在《国富论》第一卷第十章（1776）中写道，保险费“必须足以弥补通常的损失，支付管理费用，并提供一份同额资本在任何通常的贸易中所能获得的相等的利润”。

1. 公平合理原则

公平合理包括了保险人与投保人之间的公平合理以及各投保人之间的公平合理。保险人与投保人之间的公平合理又指保险的定价公平合理，既要保证保险人必要的资本收益又要保证投保人的利益不能受到损害。各投保人之间的公平合理则是指根据风险大小的不同对被保险人按照不同的费率收取。

1. 相对稳定原则

相对稳定是指保险费率的相对稳定，不能频繁调整，否则让被保险人无所适从，也使得保险公司的财务核算系统频繁更改造成混乱。当然，稳定是相对的。随着社会的发展、事物的变化，对于不适当的费率，应当根据实际统计资料加以整理，以符合实际情况。

1. 正向激励原则

正向激励原则是指保险费率的厘定要能够约束被保险人的行为，能促进防灾防损，减少保险事故，起到正向激励的效果。

### 2.1.3 车险费率厘定中关键指标—索赔频率与索赔次数

在上文中提到的保费三个组成部分中，未来的期望赔款是最为关键的因素，即纯保费，因此本文着重讨论纯保费的大小与风险因子的关系。

纯费率是指保险公司对每一风险单位的平均赔款金额，通常用赔款总额与风险单位数之比进行估计，其计算公式如下：

，其中P表示纯费率；L表示赔款总额；E表示风险单位数。

如果用N表示索赔次数，则纯费率也可以表示为：



N/E：是索赔次数与风险单位数之比，表示每个风险单位的索赔次数，即索赔频率。

L/N：是赔款总额与索赔次数之比，表示每次索赔的赔款金额，即索赔强度。因此，索赔频率和索赔强度是费率厘定过程中所要研究的重要指标之一。

## 2.2 车险费率厘定的基本方法分析

在实际保险定价中，根据整体定价思路的不同，定价方法可以分为总平均费率厘定法、分类风险费率厘定法以及个体风险费率厘定法三类。在对用户风险区分程度方面，这三种方法逐级深入。

### 2.2.1 总平均费率厘定法

总平均费率是指平均每个风险单位的保费水平。在获得经验期的风险单位数、保费、赔款和费用等数据并进行调整以后，即可厘定总平均费率。

总平均费率是纯费率和承保费用率以及利润附加率之和，其基本思想是在理赔成本的基础上附加各种必要的费用和利润得到保费。纯保费法的费率厘定可通过基本保险方程推导得出。其基本公式如下：

保费=纯保费+承保费用附加+承保利润附加

假定：R表示总平均费率；L表示根据历史经验数据得出的赔付率，即总陪付金额与保费之比；F表示根据历史经验数据得出的费用率，即除赔付成本之外的总支出与保费之比；Q表示预期得到的利润率附加。

则上述保险方程可表示为：

R=L+F+Q

纯保费法的优点是简便易行且对利润率的把控力较强，可以达到保险收费的经营目标。但费率的厘定是根据历史总体数据计算得出，没有进行个体风险的区分，缺乏对于用户风险的评估，不符合上述保险定价原则中在保险人之间的公平合理原则，容易导致较大的逆向选择风险。

### 2.2.2 分类风险费率厘定法

（1）单变量分析法

单因素分析法是每次仅仅计算一个分类变量的不同水平所对应的相对费率，通过排除其他风险变量的影响，判断单一潜在的风险分级变量是否对投保人的风险水平具有显著影响。例如，如果仅仅考虑车型和地区这两个变量，那么在根据但变量分析法厘定分类费率时，可以先计算车型的相对费率，然后再计算地区的相对费率。在分析时，索赔频率和索赔强度均可以作风险大小的衡量指标。

由于车险出险的概率分布是不均衡的（出险样本仅占所有样本的5%左右），仅用单因素对样本数据进行风险区分没有剔除选定指标之外其他因素对样本的影响，易产生偏差。

（2）迭代运算法

迭代运算法可以同时确定两个或两个以上分类变量的相对费率，是在19世纪60年代发展起来的一种普遍使用的分类费率厘定方法，最早称作最小偏差法这种方法在后来有了各种变形和发展，形成了一类特定的分类费率厘定方法。该方法的基本原理是：通过一个方程组建立风险损失数据与费率厘定因子的关系，通过迭代法求解出模型未知参数的最优解。在迭代法中，首先需要确定迭代公式。最基本的迭代公式是根据边际总和法（平衡法）、最小卡方法、最小二乘法和极大似然法建立的。不同的迭代公式具有不同的含义和性质。而各个费率因子的连接可以是加法、乘法或者是两者的混合模型。

（3）广义线性模型

尽管迭代法比单变量分析法有了很大进步，但仍然不能提供一个完整的统计分析框架。20世纪80年代，最早由Nelder和Wedderbum提出将广义线性模型应用于财产保险费率的厘定中，之后McCullagh和Nelder对广义线性模型进行了全面的总结，并将其应用于一组汽车保险损失数据的分析中。广义线性模型的建立，极大地推动了以统计方法为基石的精算学的发展。

广义线性模型是一般线性模型的推广。在古典线性回归模型中，解释变量通过线性相加关系直接影响因变量本身，并且假设因变量服从正态分布且方差为常数。而在广义线性模型中，解释变量通过线性相加关系对因变量期望值的某种变换产生影响，并且因变量不仅可以服从高斯分布，还可以是泊松分布、逆高斯分布、伽马分布等其他指数分布，且不为等方差。

广义线性模型包含随机成分、系统成分和连接函数等三个部分。各部分描如下：

1. 随机成分

随机成分是指因变量Y的概率分布。Y的每个观察值Yi相互独立且服从指数分布族中的一个分布，如正态分布、泊松分布、逆高斯分布、伽马分布等。可表示为：

f（yi/θi，φ）=

其中，i表示第i个观察值；a（φ），b（θi），c（yi，φ）为已知函数。

1. 系统成分

系统成分是指解释变量的线性组合，可表示为：η=β1\*x1+……+βp\*xp。系统成分与一般线性模型没有区别。

1. 连接函数

连接函数g单调且可导，它建立了随机成分与系统成分之间的关系，即g（μ）=η，η=g-1（μ），连接函数一般为对数函数连接、幂函数连接。

综上所述，在广义线性模型中，对因变量的预测值并不直接等于解释变量的线性组合，而是该线性组合的一个函数变换。

### 2.2.3 个体风险费率厘定法

个体风险费率厘定的基本原理是保险费率基于不同个体的风险损失来厘定的。在个体风险费率厘定中，个体风险费率厘定方法通常有两种：前瞻法和追溯法。

1. 前瞻法

前瞻法是指根据个体过去损失经验来厘定未来保险期间的费率法，具体包括经验费率法，具体包括经验费率法、表定律法和奖惩系统法。

1. 经验费率法

经验费率法是指根据被保险个体风险的损失经验计算个体风险的费率方法。经验费率法在厘定费率时要考虑投保人的理赔经验。常见的经验费率法包括两类：一类是在保险年度开始前，根据被保险人近几个保险年度的理赔经验确定下一个保险年度的续保保费；另一类是在保险年度末，根据被保险人的当前理赔经验来调整当前已缴纳的保险费。

1. 表定费率法

表定费率法是指在基础费率基础上通过改变一定比例方式来厘定车险费率的一种费率厘定方法。在实际的车险费率厘定过程中，某些个体的一些特征对费率厘定有着较大影响，而损失经验数据并不包含这些特征，此种情形下，可以在经验费率厘定之前或之后使用表定费率法，通过一定的调整比例来进行费率厘定。

1. 奖惩系统法

奖惩系统是指保险公司为了区分被保险人不同风险水平而采用的一种奖励和惩罚机制方法，又称为无赔款优待系数。奖惩系统法本质上属于经验费率法的一种，是根据被保险人上一年度的理赔情况来调整保险费率。若被保险人在上一年度无赔款，则在续保时可以享受无赔款减收优惠，降低其续保保费；若被保险人上一年度出险理赔，则根据被保险人出险的理赔次数，在续保时给予相对应的惩罚系数，适当提高其续约保费。基于奖惩系统法厘定保费不仅能够使得更加公平合理，也会激励被保险人驾车更加小心谨慎，降低其风险水平。另一方面在一定程度上抵制了骗保理赔。

1. 追溯法

是非寿险厘定方法的一种,它是基于保单当期的索赔经验对保单当期进行调整的一种费率一定方法。费率调整会受到最高追溯保费和最低追溯保费的约束。在实际费率厘定中,被保险人在投保年度初期预缴纳一定车险保费,在保险期末时根据被保险人的实际损失经验来追溯车险保费。在一定程度上,有助于鼓励投保人主动控制驾驶风险,降低其保费水平。

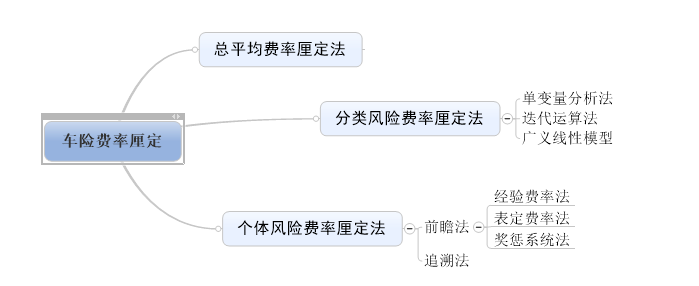


图2.1 费率厘定方法结构图

## 2.3车险费率厘定的风险因子选择

为了使得车险费率厘定更加公平、积极、有效，就必须对被保险人进行风险区分。而影响被保险人出险的风险因子众多，同时每种因素对于赔付成本的影响程度也不尽相同。整体来看，风险因子可以分为两大类：一类是与车辆相关的风险因子，包括厂牌车型、购置价格、使用性质等；另一类是与驾驶员自身情况相关的风险因子，包括驾驶员的年龄、性别、婚姻状况、驾龄、居住状况、家中私家车数量等。因此，车险费率厘定通常可分为从车费率模式和从人费率模式两种类型。

### 2.3.1 从车静态费率模式

从车费率模式是指在确定车险费率过程中以机动车辆的风险因子作为影响费率因素的模式。从车费率模式主要考虑的风险因子有：

（1）车辆使用性质

车辆使用性质不同，其面临的安全风险也不同，具体可以分为营运用车和非营运用车。相比非营运车辆而言，营运车辆的使用频率高、风险暴露高，事故率也就更高，因而其面临安全风险更大。在营运车辆中风险大小的排列顺序是：租赁车辆 > 出租车辆 > 公路客运车辆 > 公交客运车辆。

（2）车辆类型

车辆类型是指车辆的一种形式，根据车辆的使用功能可将机动车辆分为客车、货车、专用车、摩托车和拖拉机等五种类型。客车主要用来运送人员，风险主要以运客能力来衡量，即座位数，它直接关系到乘客责任风险和第三者责任的风险；货车主要用来运送货物，风险主要以货运能力来衡量，即吨位数；特种车如油罐车、冷藏车、工程车、起重车、救护车等，根据车辆不同的使用性能来衡量其特殊的风险性；摩托车主要考察其功率的大小；拖拉机的风险大小主要考察驾驶员的技术水平。

（3）排气量

排气量是用来衡量发动机规格的一个参数，它体现了汽车的动力性大小，即排气量越高其动力性越大，对于同一类型车而言也就意味着最高速度越快，因而风险也就越高。

（4）车龄

车龄是指最初车辆购置之日至投保之日止期间的年限。车辆状况同车龄有着直接的关系。一般来说，车龄越高其磨损与老化程度越高，从而导致车况越差，因而风险也就越高。

图2.2 故障率与用车强度（时间或行驶里程）之间的关系

（5）购置价格

车辆的维修成本与其购置价格密不可分，对于盗抢险尤为明显。

（6）行驶区域

车辆行驶区域是指车辆行驶的地域范围。根据我国区域地理情况不同将车辆行驶区域分为省内行驶、国内行驶以及出入境行驶三类。由于车辆行驶范围不同，驾驶员对不同地区的交通规则、地形、地貌等熟悉程度不同，以及在不同区域造成损失承担的赔偿责任不同，风险状况也不同。一般而言，风险程度随着行驶区域的扩大而增加，即出入境行驶风险>国内行驶>省内行驶风险。

（7）厂牌车型

由于世界各国汽车制造厂众多，不同的厂家生产的车辆的特点不同。从很多描述车辆的文献或是杂志中，我们都可以获知：美国和西北欧车辆首先非常重视的是安全性；日本车的性能较好，但在安全性方面不如美国和西北欧。整体而言，厂牌车型的风险排列状况为：美国、西北欧车<日本车<韩国车<国产车<其他类车。

除了上述风险因素，各个保险公司基于自身的实际业务需求，在车险费率厘定时还考虑车辆的其他风险因子，如车辆的安全状况、车身颜色等。

### 2.3.2 从人静态费率模式

从人费率模式是指在确定保险费率的过程中以驾驶员的风险因子作为影响费率因素的模式。随着车联网的发展，目前通过OBD等车联网设备可以获取到车主实时的驾驶行为，从而车主驾驶行又为可以细分为静态因素和动态因素（详见下一节）。

从人费因素主要考虑如下风险：

1. 年龄

诸多文献研究表明，机动车辆事故率与驾驶员年龄有着很高的相关性。一般而言，年轻人因年轻气盛、偏好冒险、驾驶经验不足等原因，肇事率最高；中年人驾驶经验较为丰富、处事冷静沉稳，肇事率较低；老年人因健康状况，视力开始下降，遇事反应较慢，肇事率也比较高。

1. 性别

交通事故率与性别有着密切的关系。一般而言，女性驾车比男性驾车更加谨慎。男性的性别特征决定了其驾驶行为更具有冒险性，驾驶整体速度较快。同时，在饮酒肇事事故中男性的比例明显高于女性。因而男性驾驶员在重大事故肇事概率较女性驾驶员更高。

1. 职业

由于职业的不同，人们的生活方式与性格特点往往也有着不同程度的差异，这些因素会导致肇事率的不同。统计资料表明，白领肇事率要比非白领驾驶员低；飞机驾驶员、流行音乐家事故率要高于教师；重体力劳动者事故率高于内勤行政人员。除此之外，不同车型的零整比也不尽形同。在购置价相同的条件下，零整比系数越高，赔付成本越高，相应保费就可能越高。

1. 婚姻状况

研究表明，已婚驾驶员的肇事率低于未婚驾驶员肇事率。这是由于已婚驾驶员有家庭责任和家人的督促会小心驾驶，因此其肇事率较低；而未婚驾驶员，由于无牵无挂，驾驶车辆的稳定性不如已婚者，其肇事率较高。

1. 驾龄

驾龄可以作为驾驶经验的衡量指标，它与出险率有非常明显的相关关系。通常驾龄长的司机事故概率低，驾龄短的司机事故率高。据统计，全国范围内，不满3年的驾驶员肇事占肇事总数的三成以上；在美国，驾龄不满2年的驾驶员肇事占肇事总数的比例约为12%，而此类驾驶员人数仅占持照驾驶员人数的5%。由此可见，以驾龄的长短来衡量驾驶员的风险大小是合理的。

1. 驾驶员的学历

一般来说，驾驶员的文化素质高低与事故发生率成负相关。高学历者驾驶风格稳重、低速、让他人先行；反之，低学历者驾驶风格粗鲁、喜欢高速、强行占道多。所以，高学历的驾驶员的风险比地学历的驾驶员风险小。

1. 犯罪记录

犯罪记录在一定程度上可以反映出被保险人的道德水平以及遵守法律法规的自我约束能力。若驾驶员不能有效约束自己遵守交通规则，注意避让行人，则该驾驶员出险的风险较高。

除了上述因素外，在实际车险费率厘定中还有驾驶员的其他风险因子，如驾驶员的心理状况、性格、吸烟酗酒等。

### 2.3.3 从人动态费率模式

随着车联网技术的发展，安装在车辆上的车联网终端设备可以提取到车辆的实时信息，从而可以获取到驾驶员的动态驾驶数据。

从制度层面看，随着世界各国车险费率市场化进程的不断推进，各个保险公司的自主定价权范围也在不断扩大，这为依据车联网设备提取到的动态数据定价提供了政策层面的支持。从各国车险费率市场化的进程来看，美国从20世纪60年代开始进行了保险条款费率市场化改革，实现真正意义上的独立费率厘定和自主化经营。

David.A.Leaveah 指出先前的保险费率监管模式使得保险公司容易处于被动地位，面临各种风险的同时还需要遵守限定性条款费率统一费率，逆向选择和道德风险问题突出，使得保险公司的经营风险非常严峻。欧洲从20世纪90年代开始进行保险条款费率的市场化改革，引入竞争机制，提高保险行业的市场效率。Jean Lemaire介绍了欧洲保险监管模式从条款费率的严格监管到放松管制的变迁，继而强化监管保险公司的偿付能力，并以比利时为例，具体介绍了车险条款费率市场化的过程及其多方面的效应。在亚洲市场上，韩国从1994年逐步实现车险费率自由化，从而提升了保险公司的竞争力，促进了保险产品的多元化。李淳正指出在费率自由化之前，车险过分强调了保障受害人的社会利益性，忽视了反映保险特征的产品和费率，从而导致市场上的保险产品基本相同。随着费率市场化的到来，各保险公司纷纷推出了各自独特的保险产品，增强了保险市场的竞争力，也推动了保险市场的发展。

从动态行为因素的合理性来看，20世纪六十年代开始，驾驶行为模型的研究逐渐引起人们的关注。下面对驾驶行为理论做一个简单的梳理。

以按时间划分来看，这些模型大致可以分为两类：一类是研究当前驾驶行为的影响因素，即“当前”驾驶行为的研究；另一类是研究预测驾驶员未来驾驶行为的方法，即“未来”驾驶行为的预测。

对于“当前”驾驶行为的研究，目前的理论模型已经能够较好地给出分析和解释。这些理论模型在认识驾驶行为方面主要解决了以下四个问题。

1. 驾驶行为是什么？——描述性模型

描述性模型主要有层级模型和异常驾驶行为模型两种。层级模型将驾驶中的处理过程按照时间长短和认知的先后顺序划分为3部分：计划层、策略层和控制层。1990年里森(Reason)提出了事故原因模型(Accident Causational Model, ACM) ，他认为事故的发生只有在违章、技术失效和失误等因素在时间轴上重合时才会发生，ACM模型的建立奠定了异常驾驶行为模型的基础。在事故原因模型中，所有导致事故发生的因素被分成4个层面：不安全行为（如违章和失误）、行为的前提条件、组织策略、不安全监督。基于该模型理论，理森还开发了驾驶员行为问卷(DBQ,Driver Behavior Questionaire)，被广泛地应用于调查不同地区的驾驶员的异常驾驶行为。在我国，Xie和Parker（2002）在理森驾驶员行为问卷的基础上发展出了适应我国驾驶行为的调查问卷CDQ(Chinese Driving Questionnaire), 并得出了不同的代表因子。

2.驾驶行为如何决策？——信息处理模型

Wickens（1992）提出的有限能力信息处理理论用来广泛解释人类的一般信息处理过程。Shinar（2007）将此理论延伸到了驾驶领域，提出了一个有限能力的驾驶员信息处理模型，用来描述驾驶员在驾驶过程中信息处理的全过程。与此同时，我国学者贾鸿飞(2006)也在对由人、车、路、环境所构成的动态交通系统建模与分析的基础上，提出了驾驶员的信息处理结构模型。

3.驾驶行为为何如此决策？——动机模型

信息处理模型虽然解释了驾驶员对于外在信息的反应，但这仅仅是一种被动的反应，并非主观意愿，它没有解释为何会做出某种驾驶行为决策，而动机模型弥补了这一空白。动机模型阐述了驾驶员在驾驶过程中会在一定程度内主动控制他们所遇到的风险,但这种控制并非总是向着降低风险的方向。具体来看，驾驶员在驾驶过程中不仅仅有保障安全的动机，同时还有希望节省精力、缩短时间、展示名誉地位、获得舒适感等多种与驾驶安全相悖的动机，目前此理论仅为定性。

4.环境如何影响驾驶行为？——情境中介理论

情境中介理论强调了人的因素以及人格中的倾向性在事故卷入中的作用,指出人格可以和其他因素产生交互作用,如情境因素。

除了研究“现在”时点发生的驾驶行为，还有一类驾驶行为理论模型，这些模型通过研究过去及当前的驾驶行为或其他一些因素，再利用现有信息预测“未来”可能发生的驾驶行为。

1.理性模型与感觉模型

理性模型与感觉模型解释了驾驶员的驾驶行为形成机制，即驾驶行为是如何形成的，是由哪些因素影响并决定的。其中，阿耶兹（Ajzen）提出的计划行为理论假设人是完全理性的，从理性的角度分析了驾驶行为形成的决定性因素。而达马西奥（Damasio1994）所提出的躯体标识理论则认为感觉也是行为的决定因素之一，进而从感觉的角度分析了驾驶行为是如何形成的。

2.安全驾驶行为模型

费许尔（Verschuur）和哈特（Hurts）将里森的事故原因模型与阿耶兹的计划行为理论相结合得出了安全驾驶行为模型。该理论认为不安全的驾驶行为可以直接导致事故的发生，其中不安全行为包括疏忽失误、危险失误、违章等。除此之外，该模型还增加了致因行为（causal behaviors） 因素，又可以分为计划行为理论变量与策略变量。

3．威慑理论

威慑理论又被称为惩罚规避，它从社会学和心理学的角度解释了法律惩戒对于人们驾驶行为的影响。该理论被广泛地应用于驾驶安全的评估与道路安全法律法规的制定中，例如对于危险超车、超速以及酒后驾驶行为的惩戒。



图2.3 驾驶行为理论结构图

下面针对车联网终端设备可以采集到的从人动态数据进行简单分析。

1. 年度行驶里程数

年度行驶里程反应了车辆使用强度，年行驶里程越高，风险暴露也就越大，因而风险也随之上升。

1. 平均行驶速度

平均行驶速度在很大程度上表现了驾驶员使用车辆的习惯以及车辆运转的状况，在同一地区使用车辆的平均速度越大，相应的风险也就越大，这是毋庸置疑的。但在实际生活中，市区内行驶的平均时速和高速公路行驶的平均时速相比，必然是后者更高，但是就单位公里而言，高速路行驶的单位公里风险更小，所以这就会导致一种不合理的现象：主要在高速路行驶的车辆平均行驶速度高，而相应的实际风险较小，却在平均行驶速度这一因素上被评估为更高风险级别的车辆。为了解决这一问题，就需要车联网设备搜集更多的车辆行驶区域数据来对平均速度进行划分，针对同一区域内，平均速度越高，风险越大，当然这也给数据搜集和处理带来了更大的难度。

1. 每百公里急加速/减速次数

车联网数据采集的另一个新的风险数据是急加速/减速情况。急加速/减速情况是比较新颖的风险评估因素，某些车联网设备会记录下急加速/减速前15秒和后30秒的记录，并且优先保留这类数据，这些数据与事故的发生有着高度相关性，可以在很大程度上反应驾驶员的驾驶习惯。急加速/减速的次数越多风险越大，反之则风险越小。

急加速/减速的次数与驾驶者的驾驶习惯相关，换言之与驾驶员的个性、态度、经验等因素相关，而根据信息处理模型所得到的结论，驾驶员的个性、态度、经验等因素会通过影响驾驶员的认知和注意能力、决策以及反应能力等，从而最终决定驾驶行为。因此，采用急加速/减速次数作为风险定价因子将使得车险的定价更加准确。

1. 平均转向速度

这一数据是从平均速度中分离出来的，车联网车载设备会记录转向盘超过一定角度时行驶的平均速度。这一因素与平均速度相比更加能够真实地反应驾驶员的行为习惯，快速转向的驾驶员一般具有更强的侥幸心理，风险意识比较差，而习惯低速转向的驾驶员，其安全意识更强。

与急加速/减速次数相类似，平均转向速度也取决于驾驶员的个性、态度、经验等因素，将会对驾驶行为乃至车险定价有显著影响。

1. 每次出行时段

具有相同里程数的车辆风险不一定相同，车辆行驶的时间段不同，面临的风险差别很大。上下班高峰时段的行车风险大于非高峰时段的风险，同样节假日出行的风险也高于非节假日。

出行的时段通常与出行动机密切相关，基于不同的出行动机也会产生不一样的风险驾驶行为。

1. 每百公里最佳车速驾驶里程

一般来说汽车行程的最佳时速是30~70公里（高速除外），如果低于这个时速，则很可能正在堵车，车流量密集，车辆之间的剐蹭等小事故发生概率高；而高于这个时速，在城市道路上就会显得行驶过快，遇到突发事件的时候难免会措手不及；而在最佳时速区间里驾驶的安全系数最高。

1. 发动机转速

在车辆行驶过程中发动机扮演者很重要的角色，在车辆行驶中记录下发动机的转速对于监控车辆风险有很重要的作用。发动机的转速越高，说明驾驶者对车辆使用的程度就越深，相应的风险就越大，反之则越小。

1. 平均油耗

油耗是行车的一个重要指标，表明了车辆的耗能。不同型号的车辆具有不同的标准油耗，为了了解驾驶员的行车习惯，可以考察车辆平均油耗与标准油耗（官方公布的城市最低油耗标准）的比值，比值大于1则说明驾驶者的驾驶习惯更激进，小于1则意味着更谨慎或者技术更娴熟，此比值应与风险呈正相关。

1. 居住地环境

投保人的住所环境是车辆使用环境的一部分，它主要确定车辆停泊时的风险。大部分的车辆在大部分的时间中都处于非工作状态，因此停车场所的环境对风险有很大影响，主要会影响车辆的破盗风险、划痕风险、意外破损风险等。

1. 每百公里夜间行车占比

夜间行车的风险远远大于日间行车的风险，所以夜间行车比例越大就意味着车辆风险越高，相应的保费就应越高。

1. 温度、湿度

温度和湿度能够描述车辆的行车环境，温度过低、湿度过大说明可能在雨天或雪天行车，地面附着力低，风险更大。

1. 车辆悬挂压力

车辆的悬挂装置压力能够有效地反应车辆的行驶路况，悬挂压力冲击大，则说明车辆的行驶路面不平整，风险越大，例如，山路会使得车辆的悬挂不断地受到冲击，而高速公路则使车辆的悬挂压力更稳定。

1. 每百公里山路行车里程

山路是一类比较特殊的道路，弯道多、车道窄、视线不好，突发情况较多，是事故的高发地区。通过车联网可以在卫星远程定位地图上预先设定某地区为山地区域，当汽车驶入该区域时就自动统计车辆的山路行驶里程，作为积累数据。

1. 疲劳驾驶

疲劳驾驶，是指驾驶人在长时间连续行车后，产生生理、心理机能失调，而开始出现驾驶技能下降的现象。疲劳驾驶会影响到驾驶人的感觉、知觉、注意、思维、意志、决定、判断和运动等诸方面，从而增加肇事风险。

从风险因子相关度分析角度来看，信息处理理论讨论了驾驶人本身静态特征对驾驶行为的影响，包括态度、经验、个性、实力、损伤等会通过影响驾驶人的认知、反应、决策等方式最终决定驾驶行为。而动机理论认为驾驶员的行为调整取决于其可接受风险程度的大小，当驾驶员面临的外部风险程度不同时，驾驶员会采取不同的驾驶行为以达成自己的动机。这些动机常表现于行车速度、交通违规等方面，并受到驾驶路线、行车时间等因子影响。因此，在拥有驾驶行为提取设备后后，在保险定价因子中加入驾驶时间、车速、路线等相关因素是合理的。

各个行为指标因子可以用相关的驾驶行为理论来解释，具体如下表所示。

表3.1 车险定价因子与相关驾驶行为理论

|  |  |  |
| --- | --- | --- |
|  | **车险定价因子** | **相关驾驶行为理论** |
| **传统车险的从人因素** | 性别 | 描述性驾驶理论 |
| 年龄 |
| 驾龄 |
| 职业 | 信息处理理论 |
| 婚姻状况 |
| 信用记录 |
| 犯罪记录 |
| 吸烟酗酒习惯 |
| 交通违规情况 | 动机理论 |
| 住所环境 | 情境中介理论 |
| **动态从人因素** | 年度行驶里程数 | 描述性驾驶理论 |
| 每百公里急加速/减速次数 | 信息处理理论 |
| 平均转向速度 |
| 每百公里最佳车速驾驶里程 | 信息处理理论/动机理论 |
| 平均行驶速度 | 动机理论 |
| 每次出行时段 |
| 居住地环境 | 情境中介理论 |
| 每百公里夜间行车占比 |
| 每百公里山路行车里程 |
| 车辆悬挂压力 |

# 第三章 P2P平台商业车险特征分析及厘定模型

## 3.1 P2P租车平台商业车险特征分析

### 3.1.1 我国当前商业车险费率厘定方法

目前，我国乘用车车险主要分为交强险（强制性购买）和商业险（资源购买）。商业车险主要包括：车辆损失险、商业第三者责任险、全车盗抢险、车上人员责任险、不计免赔特约条款、玻璃单独破碎险、车身划痕损失险、涉水行驶损失险等。其中，车损险和第三者责任险是车主所购买的最主要部分。

交强险的保费是固定的950元/年。商业车险基本只针对车辆购置价格（即投保金额）、车龄所属档次、车辆用途以及上年驾车出险情况进行用户区分。

以车损险为例，具体保费计算步骤：第一，判断车辆购置价格作为投保金额。第二，根据车龄、车辆用途查询所属区间保险费率。第三，根据保费=基础保费+保险金额\*费率这一等式算出车损险的大小。

再以第三者责任险为例，具体保费计算步骤：第一，用户选择三者投保金额，第二，根据保费=保险金额\*费率这一等式算出第三者责任险的大小。

最后，每家保险公司根据续保人上年的理赔记录还会给予投保人一定的优惠折扣。

### 3.1.2 我国当前商业车险费率市场化改革进程

面对外部经营环境的改变和保险公司内部竞争力提升的压力，现行的商业车险条款费率已不能满足汽车用户日益提高的保险需求，保险公司也存在承保理赔服务不到位等诸多问题。车险费率进一步细分是车险健康发展的必然趋势，车险条款费率改革实在必行。2012年3月8日，保监会发布《关于加强机动车辆商业保险条款费率管理的通知》，要求进一步完善商业车险条款费率管理制度，逐步建立市场化导向的条款费率形成机制，这是车险费率市场化改革正式启动的信号。这说明，我国车险市场正在经历由最初的统一费率到差别费率的转变。

在即将实施的新一轮车险费率市场化中，所有的改变几乎都直接体现在了保费的计算上。

保费=[基准纯风险保费/(1-附加费用率)]\*费率调整系数

对于第一项“基准纯风险保费部分”，车险费率改革的新思路较之前已有明显调整，最大的变化是：在以新车购置价、车龄、座位数等因素为核心的现有车险定价模式基础上，实时国际上同行的车型定价模式，最终实现品牌车型的差异化定价。不同车型的零整比就会影响到车损险的定价，零整比系数越高，赔付成本越高，相应保费就可能越高。

公式中第二个变量“附加费用率”，其数值则选择平移2012-2013年行业费用率的均值水平35%。新的保费公式下，附加费用率正式成为保险公司自负盈亏的根本原因。这一指标可以理解为业务获取成本与公司的运营成本。

而针对公式的第三大变量“费率调整系数”，保监会为其明确了三类相关因子系数，分别是“无赔款优待系数”（NCD）、“自主核保系数”以及“渠道系数”。在费率调整系数中，新费改中引入了自主核保系数，删除了旧体系中处理NCD之外的13个习俗。“自主核保系数”的出险，给市场留下自主核保的空间，好业务与坏业务的界限更为模糊，因此对公司风险筛选能力的要求更高。

总体来看，我国目前乘用车商业车险用户区分因子较少，并且其中大部分因子主要是从车因素。定价主要还停留在总平均费率厘定法阶段，个体风险的区分较少。

### 3.1.3 P2P商业车险特征与厘定方法分析

租车公司是指给租车人提供租车服务的公司，租车服务可以满足租客们节约养车成本以及在时间、地点、车型品牌等方面灵活选择的需求。近些年租车业务量快速增长，同时随着移动互联网的飞速发展信息和实物之间、线上和线下之间的联系变得愈加紧密，使得用户可以随时定位搜索周边租车/用车信息，除了传统的重资产型（租用车辆属于租车公司财产）租车公司，P2P租车公司开始出现。

P2P租车公司本质上是一个租车中介，它将车主和租客的信息有效地整合在一起并提供第三方监管保障服务。P2P租车可以有效整合闲置车辆资源，车主可以在不用车时将车共享出租，获得收入；对于租车来说，打开手机APP就可以搜索到附近可租车辆，方便快捷，无需专程跑到租车网点取车。

对于P2P租车模式来说，风控体系的建设是至关重要的。为了防止丢车问题的发生，租车公司（包括P2P租车公司）一般都会在其平台上的车辆安装车联网盒子设备，可以实时定位车辆，获取到汽车的行驶状态等信息。同时，为了避免车主与租车出险纠纷的发生，除了车主必上的交强险以及商业车险之外，租客也需再购买购一份商业车险。

P2P平台的商业车险是传统车险的衍生品，与传统乘用车保险有几点不同，具体可以分为：保险参与人关系的变化、保险标的的变化、被保险车辆性质的变化以及保险周期的变化。具体分析如下：

1. 保险参与人关系的变化

传统乘用车商业保险的投保人是车主，而P2P平台的商业车险投保人是租客。由于这一性质，例如车龄等反向激励因子就不应纳入到被保险人风险选择框架中。（传统车险车龄越长，保费越高；而一般车主愿意延长汽车使用寿命，但租车平台上，用户更愿租新车。）

1. 保险标的的变化

传统乘用车商业保险根据标的的不同又可以分为车辆损失险、商业第三者责任险、全车盗抢险、车上人员责任险、不计免费特约条款、玻璃单独破碎险、车身划痕损失险、涉水行驶损失险等。而目前P2P平台所出售的商业车险是混合标的，即将上述保险产品打包再按赔偿限额和免赔额不同来出售。

1. 被保险车辆性质的变化

P2P平台上所提供的车辆在车主自用时属于乘用车范畴，而一旦租给租客就变成了商用车的性质。而商用车有着易于标准化的特点，OBD等车联网设备也最先在商用车上普及。目前国内的租车平台基本已经实现了给每辆车安装汽车数据采集设备。这样一来，租客的驾驶行为就能够被记录下来。

1. 保险周期的变化

传统商业车险一般的保险周期是一年，而P2P平台商业车险的周期为一天。这种短期保险的特点可以更好地将驾驶行为定位到同一个驾驶员（租客）身上。

以宝驾租车为例，目前P2P租车平台商业车险费率的厘定仅基于车辆的租金大小（与车辆的购置价格高度相关）以及被保险人的驾龄是否已过了实习期。具体厘定方法如下如下：

基本保费=租金\*10%（各租车平台的费率大小不同，且根据保险金额的高低也有所调整）

不计免赔额保费=租金\*8%

此外，如果驾驶员仍处于实习期，则其所支付的保费是其他驾驶员购买相同保险所支付保费的两倍。

## 3.2 车联网的构成以及在保险中的作用

随着汽车智能化的提升以及互联网在各个行业中的渗透，车联网成为未来交通的发展方向之一。由于易于标准化和迫切的需求性，车联网的落地首先会在商用车上实现。P2P租车平台为了防范丢车风险以及对驾驶人风险更加及时的监测，目前都会在加入平台的车辆上安装车辆网设备（OBD等盒子产品）。

### 3.2.1 车联网构成简述

车联网本质上是汽车移动物联网，它是指利用车载电子传感装置，通过移动通信技术、汽车导航系统、智能终端设备与信息网络平台，使车与路、车与车、车与人、车与城市之间实时联网，实现信息互联互通，从而对车、人、物、路、位置等进行有效的智能监控、调度、管理的网络系统。

车联网在概念上有广义与侠义之分。从广义上讲，车联网在技术层面可以分为三大板块，包括远程信息处理（Telematics）、智能交通系统（Intelligent Transportation System，ITS）和汽车电子。而从侠义来讲，远程信息处理系统就可以看做是汽车互联网，即车联网，它只包括“一人一车”相关的部分。Telematics是借助于车载电子设备，即车载智能交互（Machine-to-Machine/Man,M2M）终端，来接收车辆的相关数据，从而随时给行车中的人们提供驾驶所需的信息。

车联网系统由“端管云”三层体系租车，分别是端系统、管系统以及云系统。端系统是智能终端传感器，用来获取车辆行驶信息以及感知行车状态；管系统用来构建车与车（V2V）、车与路（V2R）、车与人（V2H）之间的连接，实现车辆自组网及多种异构网络之间的通信；云系统是车辆运行的信息平台，它存储了大量多源信息，其应用系统是车辆数据的汇聚、计算、调度、监控、管理与应用的复合体系。

### 3.2.2 车联网设备在保险中的作用

1. 车险费率分级的作用

车险费率浮动系数中环境因素并不能够完全反映车辆真实的行驶区域。目前的费率分级方式是静态的费率分级机制，对于驾驶人行为的永久性改变，保险人难以及时发现，保险人往往是在被保险人发生事故之后做出相应调整，比较被动。车辆网技术的引入，使得对驾驶员的驾驶行为以及安全系数的量化分析成为可能。

1. 防灾防损与车联网技术

车联网技术的引入可以动态的检测驾驶员的驾驶行为，一旦发现驾驶行为出险暂时性的异常，保险人可以及时通过温馨提示等辅助措施纠正驾驶人的行为，达到防灾防损的目的，最终降低车险的赔付率，提高保险公司车险的竞争力。

1. 车联网技术在理赔中的作用

车联网技术的引入，可以使保险公司获得事故发生时的准确信息，包括事故车辆状态、现场周围环境等。通过充分运用这些信息，可以使保险公司更高效、准确地进行理赔。目前，一些保险公司陆续推出“闪赔”等服务，当定损金额在规定范围内时即可享受该服务。这种服务模式提高了理赔效率，增强了客户感受，有助于提高保险公司的竞争力，增强品牌形象。但是，这些服务在提高保险公司经济效益的同时也给保险公司带来了风险隐患。车联网技术的引入，可以为快速理赔服务提供准确的数据支持，减少为提高服务水平而带来的风险附加。

目前，P2P租车平台运用车联网设备主要目的仅是对车辆进行定位防止车辆的丢失，其功能应用的拓展还具有非常大的空间。

## 3.3 P2P商业车险厘定步骤分析

目前，车险费率改革正在加速进行中，我国车险的定价正从总平均费率厘定法转变为分类风险费率厘定法。P2P平台商业车险由于体量小，周期短，调整灵活的特点，不妨将其作为车险改革的突破口。因此笔者着重从区分用户风险大小的角度来设计车险厘定方法。

首先，从车因子方面来看。交通部数据显示，由于车辆本身的问题造成的交通事故率仅为10%左右，因此可以将车型、排量等从车因子合并（车龄具有反向激励作用）。另一方面，虽然出险概率与从车因子相关度较少，但出险金额的高低和车辆本身零部件的价格高度相关。基于这点考虑，可以将租金（租金与车辆购置价格高度相关，一般由车主决定）作为模型中的乘积因子。

其次，静态从人因素的筛选原则主要有：真实数据可得性、相关性等。目前租车平台根据租客的身份证和驾照可以准确判定的从人因素主要有性别、年龄、驾龄、地区（汽车使用地区）等。下文中笔者主要讨论如何根据从人因素来模拟出险概率，而不继续讨论与出险概率挂钩的车险费率表（这张表的制定还要根据保险运营费用等利润分析来拟定）。

再次，根据计划行为理论、躯体标识理论等前人的研究成果，我们可以通过驾驶员过去的行为及其决定因素，来预测驾驶员未来发生相同行为的可能性。租车平台恰有对于车主驾驶行为数据的获取渠道。由于是一个新开拓的部分，我们可以将这部分因素的判别放入不计免赔额费率厘定中。具体方法是通过这部分数据来给租客驾车行为进行打分，给予不同评分段的租客给予不同免赔额的费率。这样做还可以起到正向激励的作用，使租客规范驾车行为，减少出险概率的发生。下文中笔者主要讨论如何根据动态因素来给租客的驾驶行为进行打分，而不继续讨论驾驶评分与出险概率挂钩的免赔额费率表（这张表的制定还要根据保险运营费用等利润分析来拟定）。同时，限于无法获取对比样本，本文不考虑驾驶行为评分模式的正向激励作用。

P2P租车平台的车险厘定步骤的框架如下图所示：

保费=日租金\*（基础费率+不计免赔额费率）

根据所拟合风险大小，对应费率表进行查询。

根据所拟合风险大小，对应费率表进行查询。

利用熵权——层次分析法设计驾驶行为评分表。

利用基于Tweedie分布的广义线性模型进行风险拟合。

图3.1 P2P商业车险厘定框架

# 第四章 从人静态因素的风险度量

## 4.1 样本数据说明

本文的数据全部来自宝驾P2P租车平台。

首先将总体样本按地区划分进行分层，从上海、北京两个城市分别抽取500个样本。抽取样本时以订单编号（一年及一年以内的订单）为关键字，进行随机（订单时间随机）抽样。

抽样后的样本容量为1000，又进一步按租车地区将其分为两组，上海地区500个样本，北京地区500个样本。其中，北京地区样本中有18个样本为出险样本；上海地区样本中有17个样本为出险样本。

静态样本维度分别为：地区、订单编号、驾客性别、驾客年龄、驾客驾龄、实际取车时间、实际还车时间、购买保险类别、车辆品牌系列、车龄（年）、已行驶里程、出险原因、车辆损失状况、报案时间、赔付客户金额，共15个维度。

## 4.2 模型的选取

广义线性模型( Generalized linear model，简称GLM) 最早由Nelder 和Wedderburn提出，广义线性模型的建立，极大地推动了以统计方法为基石的精算学的发展，目前正逐渐运用于我国保险价格的厘定中。

广义线性模型与一般线性模型的主要有以下三个区别：第一，因变量y不再局限于正态分布，而是推广到指数族分布；第二，被解释变量y的方差不必为常数，而是表示为均值的函数形式；第三，在使用GLM建模时，经过连接函数（对数连接或指数连接）变换以后均值表示为解释变量的线性形式，即g( μ) = η = Xβ。传统的单变量分析法、最小偏差法无法对复杂保险精准数据，而这正是GLM 得以广泛应用的优势所在。因此，本文选取GLM模型作为静态从人因素的拟合方法。

## 4.3 指标的选取与数据的预处理

风险因子筛选要符合①可靠性 ②相关性 ③正向激励三个原则。可靠性是指数据来源的可靠性，即数据值的判定要有可靠的来源，不能仅依据被保险人的自述；相关性是指因变量（赔付客户金额）与所选取的风险因子有明显的相关性；正向激励是指风险因子所带来保费的增加可以抑制（至少不会促进）被保险人风险等级的上升。以下风险因子的筛选要符合这三个原则。

首先做一下样本的清洗，将订单编号作为关键字去重。北京500个样本没有重复订单，上海500个样本有2个是重复订单（并且这两个订单为出险订单），剔除其中一个样本。因此总样本容量变为999个，其中北京500个样本，有18个出险保单；上海499个样本，有16个出险保单。

下面进行风险因子的筛选：

（1）驾客地区（a）：这个风险因子从租车信息平台易于获取、验证，平台上的车辆严格按照城市进行划分。本文样本只选取了上海、北京两座城市进行探讨。城市分为两个水平：北京地区a1=0,上海地区a2=1。

下面以风险因子分组数据的索赔强度和索赔频率来刻画该风险因子的风险大小（以下风险因子同）。

风险因子分组的索赔频率=该风险分组的出险次数/该风险分组的样本数量\*100，

风险因子分组的索赔强度=该风险分组赔付客户总金额/该风险风阻的出险次数，

则：

北京地区的索赔频率=18/500\*100=3.60次

北京地区的索赔强度=79339/18=4408元

上海地区的索赔频率=16/499\*100=3.21次

上海地区的索赔强度=14882/16=930元

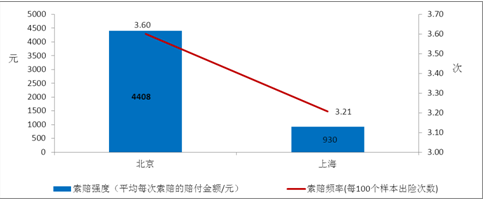


图4.1 北京、上海两地P2P商业车险索赔强度、索赔频率分布

从图中可以看出，无论是从索赔频率还是索赔强度来看，不同年龄段的风险水平具有显著的差异性。从可靠性、相关性以及风险区分的角度来看，年龄因子可以选做一个风险因子变量。

（2）驾客性别（b）：平台初始注册时会进行身份验证，因此通过身份证信息可以准确地获取到驾客性别。性别分为两个水平：性别女b1=0，性别男b2=1。

女性的索赔频率=2/137\*100=1.46次

女性的索赔强度=1726/2=863元

男性的索赔频率=32/862\*100=3.71次

男性的索赔强度=87445/32=2733元

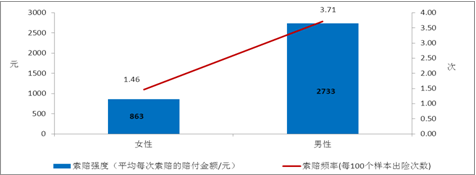


图4.2 P2P商业车险男性、女性的索赔强度、索赔频率分布

从图中可以看出，无论是从索赔频率还是索赔强度来看，男性的赔付风险要大于女性。从可靠性、相关性以及风险区分的角度来看，性别因子可以选作一个风险因子变量。

（3）驾客年龄（c）：驾客年龄同样也可以准确地通过身份证号识别出。驾驶人年龄分为六个水平：18-20岁：c1=1, 21-24岁：c2=2, 25-29岁：c3=3, 30-34岁：c4=4, 35-39岁：c5=5 ，40-49岁：c6=6；大于等于50岁：c7=7。

18-20岁之间的样本容量仅有2个，且均为未出险样本。样本容量太小，并不能说明该年龄段出险概率为0，因此剔除这两个比较极端的样本（若需要判定该年龄段的风险大小，则需扩大样本重新随机抽样）。

21-24岁的索赔频率=5/82\*100=6.10次

21-24岁的索赔强度=54028/5=10806元

25-29岁的索赔频率=16/347\*100=4.61次

25-29岁的索赔强度=14254/16=891元

30-34岁的索赔频率=6/279\*100=2.15次

30-34岁的索赔强度=12920/6=2153元

35-39岁的索赔频率=3/203\*100=1.48次

35-39岁的索赔强度=1716/3=572元

40-49岁的索赔频率=3/85\*100=3.53次

40-49岁的索赔强度=5150/3=1717元

50岁及以上的年龄段样本容量仅有1个，且为出险样本。样本容量太小，并不能说明该年龄段出险概率为100%，因此剔除这两个比较极端的样本（若需要判定该年龄段的风险大小，则需扩大样本重新随机抽样）。

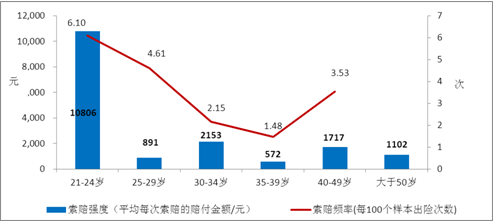


图4.3 不同年龄段P2P商业车险索赔强度、索赔频率分布

从图中可以看出，无论是从索赔频率还是索赔强度来看，不同年龄段的风险水平具有显著的差异性。从可靠性、相关性以及风险区分的角度来看，年龄因子可以选做一个风险因子变量。

（4）驾客驾龄（d）：平台初始注册时会进行驾照验证，因此通过驾照信息可以准确地获取到驾客驾龄（假设拿本当年就开车上路）。驾龄分为1年以下：d1=1，1-3年：d2=2，4-7年：d3=3，8-9年：d4=4，10年及以上d5=5。

驾龄小于1年的索赔频率=7/62\*100=11.29次

驾龄小于1年的索赔强度=7479/7=1068元

驾龄1~3年的索赔频率=12/49\*100=2.80次

驾龄1~3年的索赔强度=20134/12=1678元

驾龄4~7年的索赔频率=8/266\*100=3.01次

驾龄4~7年的索赔强度=54105/8=6763元

驾龄8~9年的索赔频率=2/85\*100=3.53次

驾龄8~9年的索赔强度=4923/3=1641元

驾龄10年及以上的索赔频率=4/157\*100=2.55次

驾龄10年及以上的索赔强度=2529/4=632元

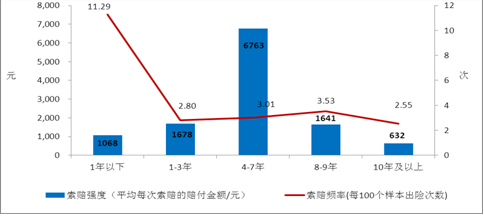


图4.4 不同驾龄段P2P商业车险索赔强度、索赔频率分布

图中可以看出，无论是从索赔频率还是索赔强度来看，不同驾龄段的风险水平具有显著的差异性。从可靠性、相关性以及风险区分的角度来看，年龄因子可以选做一个风险因子变量。

（5）从车因素：

在车主将车放在租车平台之前，会有一个车辆认证检测的环节。这个环节上可以确切获取到车辆品牌系列、车龄以及已行驶总里程数等从车因子，符合可靠性。

据美国、英国和澳大利亚的专家学者对大量事故的深入研究，得出保险事故中与人相关的原因占93%-94%，而与车相关的原因仅占8%-12%。

与此同时，我们分析了本文所采集的样本，得到车龄与赔付客户金额的相关系数为-0.0058，车龄与是否出险（未出险计作0，出险计作1）的相关系数为0.0093，相关性很低，不符合相关性，舍去。

已行驶公里数与赔付客户金额的相关系数为0.0062，已行驶公里数与是否出险（未出险计作0，出险计作1）的相关系数为0.02。已行驶公里数与是否出险（未出险计作0，出险计作1）相关系数是正向关系，即车辆损耗越多，出险概率越大，需要交的保费也越多。从租车（即被保险人）的角度来说，也喜欢选择租用损耗小（已行驶公里数低）的车辆，不符合正向激励原则，舍去。

风险因子的选择具体如表3.1所示：

表3.1 静态数据因子的取值特征及筛选依据

|  |  |  |  |
| --- | --- | --- | --- |
| **静态数据维度** | **取值特征** | **是否选取作为风险因子** | **说明** |
| 订单编号 | 320006252425554 | 否 | 作为关键字来区分不同订单 |
| 地区 | 北京 | 是 |  |
| 驾客性别 | 男 | 是 |  |
| 驾客年龄 | 29 | 是 |  |
| 驾客驾龄 | 1年0月 | 是 |  |
| 实际取车时间 | 2015-03-02 19:00:01 | 否 | 保费单价按天计算，用租车时长乘以单价即可算出保费 |
| 实际换车时间 | 2015-03-09 20:39:01 | 否 |
| 购买保险类别 | 中国人寿，本地范围，200万，20% | 否 | 不相关，且不是静态从人因素 |
| 车辆品牌系列 | 起亚狮跑 | 否 | 无车型整体数据库参照，且不是静态从人因素 |
| 车龄（年） | 6 | 否 | 测算后发现相关度低，舍去 |
| 已行驶公里 | 8万公里 | 否 | 不符合正向激励原则，且不是静态从人因素 |
| 出险原因 | 租客自己倒车不小心撞树上了 | 否 | 暂时无法标准定量化 |
| 车辆损失状况 | 车辆右后方保险杠右后方有凹陷面积大概一个半手掌 | 否 | 暂时无法标准定量化 |
| 报案时间 | 2015/3/5 | 否 | 不相关，且不是静态从人因素 |
| 赔付客户金额 | 1100 | 否 | 作为因变量 |

因此，选择驾客地区（a）、驾客性别（b）、驾客年龄（c）、驾客车龄（d）这四个从人因素作为风险因子加入模型。

在因子选择的过程中又去除了年龄为18-20岁之间的2个样本以及年龄50岁及以上的1个样本，总样本容量变为996，其中出险样本为33个。

## 4.4 损失分布的确定

由于P2P租车平台上的保单都是短期保单，因此一张保单对应的实际索赔次数最多为1，以订单ID进行区分的索赔频率无实际意义。因此仅将索赔强度（每张订单的赔付客户金额，）作为因变量进行拟合。

按照以上四个因素可以将保单持有人的风险分为2\*2\*5\*5=100个风险单元。设随机变量Yabcd表示单元（a，b，c，d）的赔付客户金额。这里a，b，c，d分别代表地区、性别、年龄、驾龄四个风险因素的各个水平（a=1,2; b=1,2; c=2,3…5;d=1,2,…,5）。

建立广义线性模型：

E（Yabc）=μabc  (1)

ηabc =Xβ=(1，xa，xb，xd)\* β, β=(β0，βa，βb，βc，βd)T (2)

Φ（μabc） =ηabc (3)，Φ为连接函数。

由于，广义线性模型中，赔付客户金额Yabcd并不服从正态分布（对称），而保险的赔付客户金额Yabcd是大于等于零，非对称性的，且在零值时峰度非常陡峭，它的形态更近似于泊松分布（Poisson）、Tweedie分布。但由于泊松分布、Tweedie分布的峰值在零点右侧，且在零点处的概率为0，因此将Yabc平行向上移动1个单位，即设加入变量Y’ abc=Yabc+1。用Y’ abc代替Yabc的值进行拟合，即给每个赔付客户金额增加1元，不影响风险因子的分布规律。

利用SPSS软件，分别用正态分布、Poisson、Tweedie分布拟合索赔额分布，连接函数为对数连接。不同分布下的拟合优度比较详见表3.2：

表3.2 不同分布下的拟合优度

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 自由度 | 总离差 | 平均离差 | 对数似然 |
| 正态分布 | 985.0 | 2269394871.0 | 2303954.2 | -8703.5 |
| 泊松分布 | 985.0 | 564048.6 | 572.6 | -283130.0 |
| Tweedie | 985.0 | 38970.0 | 39.6 | -5623.5 |

从上表我们可以看出，正态分布的总离差最大，差异明显，Tweedie分布的总离差最小，拟合效果最好，而泊松分布的总离差居中。从对数似然值来看，泊松分布的值最小，而Tweedie分布的对数似然值最大，Tweedie分布的拟合效果最好。

对于各个参数的显著性检验我们通过SPSS分析结论如下表：

1.正态分布

表3.3 正态分布的参数估计

|  |  |  |  |
| --- | --- | --- | --- |
|  | Wald卡方值 | 自由度 | P值 |
| (截距) | 0.000249 | 1 | 0.9874 |
| 地区 | 0.000259 | 1 | 0.9872 |
| 年龄分组 | 0.000047 | 1 | 0.9945 |
| 驾龄分组 | 0.074734 | 4 | 0.9993 |
| 性别 | 0.002461 | 4 |  |

2. 泊松分布

表3.4 泊松分布的参数估计

|  |  |  |  |
| --- | --- | --- | --- |
|  | Wald卡方值 | 自由度 | P值 |
| (截距) | 0.000249 | 1 | 0.9874 |
| 地区 | 0.000259 | 1 | 0.9872 |
| 年龄分组 | 0.000047 | 1 | 0.9945 |
| 驾龄分组 | 0.074734 | 4 | 0.9993 |
| 性别 | 0.002461 | 4 | 1.0000 |

3. Tweedie分布

表3.5 Tweedie分布的参数估计

|  |  |  |  |
| --- | --- | --- | --- |
|  | Wald卡方值 | 自由度 | P值 |
| (截距) | 995.128 | 1 | 0.0000 |
| 地区 | 150.160 | 1 | 0.0000 |
| 年龄分组 | 144.751 | 1 | 0.0000 |
| 驾龄分组 | 679.871 | 4 | 0.0000 |
| 性别 | 26.747 | 4 | 0.0000 |

我们可以看出，正态分布模型中，所有参数均未通过显著性检验；而泊松分布和Tweedie分布的所有参数均通过了显著性检验。

综合来看，我们选取Tweedie分布来拟合赔付客户金额。

## 4.5 模型结果及分析

用SPSS软件，以Tweedie分布拟合索赔额分布，连接函数为对数连接，得到结果如表3.6所示：

表3.6 损失强度的参数估计值

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 因素 | 因素水平 | 自由度 | 估计值 | 标准差 | 95%置信度下限 | 95%置信度上限 | 卡方值 | P值 |
| 截距 | | 1 | 3.508 | 0.1884 | 3.13900 | 3.87800 | 346.73 | 0 |
| 地区=0 | 北京 | 1 | 1.10600 | 0.09020 | 0.92900 | 1.28300 | 150.16 | 0 |
| 地区=1 | 上海 | 1 | 0.87395 | 0.00624 | 1.03771 | 0.71018 | 134.23 | 0 |
| 性别=0 | 女 | 1 | 0.53042 | 0.18230 | 0.51812 | 0.54272 | 144.751 | 0 |
| 性别=1 | 男 | 1 | 0.98473 | 0.23942 | 0.91531 | 1.05415 | 153.293 | 0 |
| 年龄分组=2 | 21~24 | 1 | 2.00300 | 0.17700 | 1.65600 | 2.35000 | 128.042 | 0 |
| 年龄分组=3 | 25~29 | 1 | 0.53000 | 0.16600 | 0.53400 | 0.52600 | 10.493 | 0.001 |
| 年龄分组=4 | 30~34 | 1 | 1.56600 | 0.16970 | 1.56630 | 1.56570 | 9.772 | 0.002 |
| 年龄分组=5 | 35~39 | 1 | 0.87630 | 0.20060 | 0.84570 | 0.90690 | 60.958 | 0 |
| 年龄分组=6 | 40~49 | 1 | 1.23842 | 0.15746 | 1.25096 | 1.22588 | 12.484 | 0 |
| 驾龄分组=1 | 小于1年 | 1 | 0.67200 | 0.20580 | 0.26900 | 1.07600 | 10.669 | 0.001 |
| 驾龄分组=2 | 1-3年 | 1 | 0.05800 | 0.15950 | -0.25500 | 0.37100 | 0.132 | 0.717 |
| 驾龄分组=3 | 4-7年 | 1 | 0.45000 | 0.17320 | 0.11000 | 0.78900 | 6.749 | 0.009 |
| 驾龄分组=4 | 8-9年 | 1 | 0.38800 | 0.20420 | -0.01200 | 0.78900 | 3.618 | 0.045 |
| 驾龄分组=5 | 10+ | 1 | 0.03240 | 0.13240 | 0.27000 | -0.00520 | 1.234 | 0 |

在地区分组中，两地区均通过置信水平为1%的显著性检验。由于北京地区的估计值要高于上海地区，北京地区索赔强度较高。这可能与北京的路况有关，如交通拥堵情况更为严重、道路复杂等有关。

在性别分组中，男性的估计值要明显高于女性，这说明男性的赔付金额总体上来说要大于女性。这可能与男性驾驶时的谨慎度比女性要低有关。

在年龄分组中，年龄在21~24岁之间（年轻人）的估计值（2.003）明显高于其他因子前的估计值，这说明此年龄段的租客风险最大，此年龄段正处于年轻气盛的阶段，胆大却未必心细，因此风险较大。30~34岁之间的参数估计值（1.566）相对来说也比较大，驾驶的平均损失相对也比较高，这个年龄段事业、家庭压力均比较大，身体体力情况、注意力集中的问题都会成为其风险较高的原因。25~29岁之间以及35~39岁之间的人群驾驶的平均损损失相对较低，处于这两个阶段的人群相对年轻，大脑、机体反应速度比较敏锐，同时相对于30~34岁年龄段的人群生活压力较小，因此出险几率也会随之降低。39岁之后随着年龄的增大，驾驶的平均损失成本开始增大，这可能由于机体反应速度、注意力集中度逐渐降低，受身体条件所限驾驶的平均损失成本开始增加。

在驾龄分组中，刚拿到驾照的人群（驾龄1年之内）的损失程度较大。当驾驶技术熟练之后，8-9年驾龄损失程度较小，之后随着车辆使用年限的增长，其平均损失程度也在呈现逐步降低的趋势。值得注意的是，驾龄在4-7年的阶段，损失程度仅次于n刚拿到驾照的人群，这可能是因为这个驾龄段的司机没有恐惧的心里障碍，但驾车经验仍有欠缺，一旦出险就是比较严重的事故，赔付金额较高。

从Tweedie分布拟合的广义线性模型来看，它与上文所进行的单因子分析结论基本一致。这说明用广义线性模型进行多因子拟合有很好的解释效果。

# 第五章 从人动态因素的风险评价

## 5.1 样本数据说明

此部分动态数据样本均取自上文中的999个样本，其中33个是出险样本的动态数据（原本有35个出险样本，但由于2015年6月之前的样本还没有开始全面跟踪租车平台上所有的汽车行驶数据，其中一个出险样本无法获取到动态数据；35个出险样本中有两条抽取重复，删去其中一个样本）。另外35个样本是在未出险的965个样本中随机抽取的35个对比样本。

原始的动态数据是由车联网设备（类似于OBD盒子）提取到的，它平均每隔30秒（间隔不稳定，有时间隔短有时长）提取一次车辆瞬时经度、瞬时维度、瞬时速度以及瞬时行驶方向。具体指标如表5.1所示

表5.1 动态数据样本示意

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 时间 | 经度 | 纬度 | 速度 | 行驶方向 |
| 2015-11-26 10:30:16 | 116.4184 | 39.90403 | 0 | 0 |
| 2015-11-26 10:32:46 | 116.4184 | 39.90403 | 0 | 0 |
| 2015-11-26 10:35:17 | 116.4184 | 39.90403 | 0 | 0 |
| 2015-11-26 10:37:17 | 116.4184 | 39.90447 | 19 | 358 |
| 2015-11-26 10:37:47 | 116.4183 | 39.90661 | 21 | 356 |
| 2015-11-26 10:37:59 | 116.4185 | 39.90689 | 24 | 74 |
| 2015-11-26 10:38:17 | 116.421 | 39.90688 | 30 | 93 |
| 2015-11-26 10:38:48 | 116.4225 | 39.90689 | 8 | 86 |
| 2015-11-26 10:39:18 | 116.423 | 39.9069 | 0 | 82 |
| 2015-11-26 10:40:48 | 116.423 | 39.9069 | 22 | 87 |
| 2015-11-26 10:41:18 | 116.4275 | 39.90705 | 25 | 87 |
| 2015-11-26 10:41:48 | 116.4288 | 39.90706 | 9 | 90 |
| 2015-11-26 10:42:18 | 116.429 | 39.90707 | 0 | 79 |

## 5.2 模型的选取

通过动态数据来进行被保险人风险大小的甄别，笔者要设计一个对驾驶行为进行评分的模型，根据评分的高低可以判断被保险人风险的大小。如何确定每个驾驶行为指标的对于评分高低影响的权重为模型建立的关键之一。目前，有关权重的确定方法有很多。根据计算权重时原始数据来源不同可以分为主观赋权法和客观赋权法。为了避免主观赋权的主观性，选择客观赋权法来进行权重的计算。客观赋权法中最为常用的方法即为熵权法，本文采取此种方法来赋权。

熵原本是热力学中的概念，申农最早将其引入信息论，现已在社会经济、工程技术等领域得到广泛的应用。熵权法的基本原理为：如果某个因子的熵Ej越小，就表明该因子的变异程度越大，提供的信息量也就越大；某因子的信息熵Ej越大，就表明该因子的变异程度越小，提供的信息量也就越小。在综合评价中熵越大，该因子所起到的作用就越小，则其权重也应越小。

第j项指标的信息熵计算公式为：

Ej=-（lnm）-1∑PijlnPij, j=1,2,…,n

其中，m为被评价对象的数目，n为评价指标数目,

pij=，如果pij=0，我们定义pij→0, lim PijlnPij=0,

dij= ① 或 dij=②

其中，fij表示第i个样本的第j个统计量，公式①适用于指标越大越好，公式②适用于指标越小越好。

## 5.3 驾驶行为指标的选取与数据的预处理

驾驶行为因子指标的筛选同样也要符合①可得性 ②相关性 ③正向激励三个原则。可得性是指驾驶行为数据是可以通过车联网设备直接或者间接获取到的；相关性是指被保险人出险的风险大小与所选取的行为因子有明显的相关性；正向激励是指驾车行为因子所带来评分的提高能够说明被保险人驾车行为安全等级的上升。以下行为因子的筛选要符合这三个原则。

下面进行行为因子的筛选。首先将行为因子分为三大类，即行驶里程及时间、超速行驶情况以及驾车行驶情况，具体指标如表5.2所示：

表5.2 驾驶行为评价指标分类

|  |  |
| --- | --- |
| **驾驶行为评价指标分类** | **驾驶行为评价指标** |
| 行驶里程及时间 | 行驶总时间 |
| 平均单日行驶时间 |
| 行驶总里程 |
| 平均单日行驶里程 |
| 夜间行驶总时间 |
| 平均单日夜间行驶时间 |
| 夜间行驶平均占比 |
| 超速行驶情况 | 非高速公路行驶每百公里超速次数 |
| 高速公路行驶每百公里超速次数 |
| 平均加速度 |
| 驾车行驶情况 | 每百公里急加速次数 |
| 每百公里急减速次数 |
| 每百公里急变速次数 |
| 每百公里急转向次数 |
| 最大连续驾驶时间 |
| 速度标准差 |
| 非高速公路行驶的平均速度 |
| 高速公路行驶的平均速度 |

由于所获取的原始行为数据只有四个维度，在刻画出最终行为因子数据之前，需要预先构建一些中间变量。

1.distance\_delta(△s)：相邻两个取样时间点之间的地理坐标距离（米）；

这个指标的计算需要根据两个坐标点（经纬度来刻画）来计算他们之间的距离？

我们这里假设地球是一个完美的球体，那么它的半径就是地球的平均半径，记为R。以0度经线为基准，根据地球表面任意两点的经纬度可以计算出这两点间的地表距离（这里忽略了地球表面地形对计算带来的误差）。

具体计算方法：点 A的经纬度为(LonA, LatA)，点B的经纬度为(LonB, LatB)，按照0度经线的基准，西经取经度负值(-Longitude)，东经取经度的正值(Longitude)，北纬取90-纬度值(90-Latitude)，南纬取90+纬度值(90+Latitude)。经过上述处理过程后，两点坐标分别为(MLonA, MLatA)和(MLonB, MLatB)，得到两点距离计算公式如下：

C = sin(MLatA)\*sin(MLatB)\*cos(MLonA-MLonB) + cos(MLatA)\*cos(MLatB)

Distance = R\*Arccos(C)\*Pi/180

这里，R和Distance单位是相同，我们采用6371004米作为半径，那么Distance就是以米为单位。

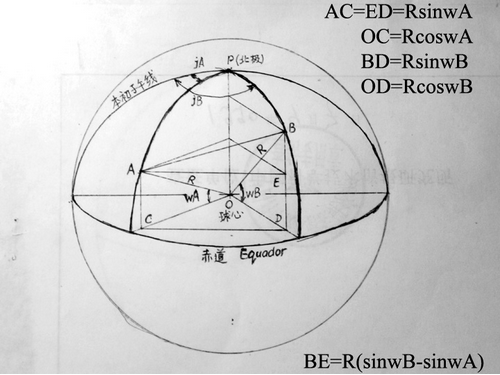


图5.1 地球上两点之间距离示意图

2. time\_delta(△t)：相邻两次记录之间的时间间隔（秒）；

这个中间变量只需用后一取样时间点的记录时间减去前一取样时间点的记录时间。

3. at\_night(f1)：判断记录时间点是否在夜间（0-否，1-是）；

若时间在20:00-06:00之间，则为1，否则为0。

4. continuous\_time(tc)：累计的连续驾驶时间(秒)

这个变量需要有一个前提判断，即判断是否是连续驾驶：连续驾驶，中间停车（瞬时速度为0）不多于两个记录时间，也就是60秒（当然也可以采用更加复杂的判断条件）；

如果判断为连续驾驶，则将此阶段连续驾驶的time\_delta(△t)累计相加得到continuous\_time(tc)。

5. acceleration（a）：实时加速度（米/秒2）。

用公式a=，

其中，V2、V1分别为相邻两次记录的时间；

T2、T1：分别为相邻两次记录的速度。

6. highWay（f2）：判断记录点是否在高速路上（0-否，1-是）；若判断在高速公路上行驶，则为1，否则为0。

是否在高速公路上行驶可以根据经纬度坐标结合地图信息来判断行驶路段信息。

但本文在写作时没有地图经纬度坐标数据库，因此采用速度来判断。具体判断为：连续5分钟超过70km/h，则判定为高速公路驾驶。

若某车并未在高速公路上行驶，连续5分钟以大于70km/h的速度行驶，则5min至少行驶了5.83km，那么这段路程中一定会驶过交叉路口，速度不可能一直保持在70km/h以上，假设不成立。那么，连续5分钟超过70km/h，则判定为高速公路驾驶,这一判断标准是可行的。

其次，基于原始数据和中间变量计算出汽车驾驶员行为评分指标。

第一类指标：行驶里程及时间

1.行驶总时间（T行）、平均单日行驶时间（t行）

T行的计算是将每段连续行驶里程的time\_delta(△t)加总求和得出。

t行的计算是用T行除以租车总时间

2.总里程（S）、平均单日行驶里程（S日）

S的计算是将每个取样点之间的距离distance\_delta(△s)加总求和得出。

S日的计算是用S除以租车总时间

3.夜间行驶总时间（T夜）、平均单日夜间行驶时间（t夜）

T夜的计算是将f1=1（判断是否在夜间）的每段连续行驶里程的continuous\_time(tc)加总得出。

t夜的计算是用T夜除以租车总时间

4.夜间行驶平均占比（W夜行）

W夜行=

第二类指标：超速行驶情况

1. 非高速公路行驶每百公里超速次数（q非）

首先，先判断非高速公路行驶超速的标准，我们以70km/s为限。

当f2=0（判断在非高速公路上行驶），v>70km/s时判断为超速，求超速和。

之后把频次数据用以每100km的里程来归一化，将求得的非高速公路超速次数除以高速公路行驶里程即得到q非。

2. 高速公路行驶每百公里超速次数（q高）

首先，先判断非高速公路行驶超速的标准，我们以120km/s为限。

当f2=1（判断在高速公路上行驶），v>120km/s时判断为超速，求超速和。

之后把频次数据用以每100km的里程来归一化，将求得的高速公路超速次数除以高速公路行驶里程即得到q高。

第三类指标：驾车行驶情况

1. 平均加速度（a均）

计算a的均值即得a均

2. 每百公里急加速次数（d加）

以丰田普拉多（2015款 2.7L 自动豪华版）为例，其百米加速为15.06s，即油门踩到最大时，达到100km/h的速度需要15.06s，也就是说最大加速度为1.84m/s2。因此，笔者把急加速定义为百米加速大于20s，即加速度超过1.38m/s2。

之后把以每100km的里程来归一化，将求得的急加速次数除以总行驶里程得到d加。

3. 每百公里急减速次数（d减）

笔者把急减速定义为-1.5m/s2,即实时加速度a小于1.5m/s2时，即定义为一次加速度。

之后把以每100km的里程来归一化，将求得的急加速次数除以总行驶里程得到d减。

4. 每百公里急变速次数（d总）

d总= d加+ d减

5. 每百公里急转向次数（d转）

求continuous\_time(tc)的最大值即可得出tmax。

6. 最大连续驾驶时间tmax

求continuous\_time(tc)的最大值即可得出tmax。

7. 速度标准差Sv均

求原始数据中每个取样时间点速度v的标准差可以得出Sv均。

8. 非高速公路行驶的平均速度（V非）

先判断是否是非高速路段行驶，f2=0，

V非=，

9. 高速公路行驶的平均速度（V高）

先判断是否是非高速路段行驶，f2=1，

V高=，

## 5.4 驾驶行为指标的再筛选与指标权重的计算

上文中在进行驾驶行为因子数据可得性时，删除了急转向次数这一指标。

根据样本数据拟合之后，通过出险样本和未出险样本的对比，我们可以判断驾驶行为因子是否均符合相关性与正向激励原则。具体结果如表5.3所示：

表5.3 样本数据对比与驾驶行为评价指标的筛选

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **驾驶行为评价指标分类** | **驾驶行为评价指标** | **出险样本结果** | **未出险样本结果** | **相关性** | **正向激励** |
| 行驶里程及时间 | 行驶总时间（h） | 13.12 | 8.65 | 是 | 是 |
| 平均单日行驶时间（h） | 3.92 | 4.81 | 是 | 否 |
| 行驶总里程（km） | 517.52 | 346.04 | 是 | 是 |
| 平均单日行驶里程（km） | 123.33 | 155.87 | 是 | 否 |
| 夜间行驶总时间（h） | 2.43 | 1.73 | 是 | 是 |
| 平均单日夜间行驶时间（h） | 0.82 | 1.00 | 是 | 否 |
| 夜间行驶平均占比 | 0.21 | 0.21 | 否 | 否 |
| 超速行驶情况 | 非高速公路行驶每百公里超速次数 | 12.78 | 18.70 | 是 | 否 |
| 高速公路行驶每百公里超速次数 | 0.83 | 1.14 | 是 | 否 |
| 平均加速度（m/s） | 0.19 | 0.20 | 否 | 否 |
| 驾车行驶情况 | 每百公里急加速次数 | 1.67 | 1.95 | 是 | 否 |
| 每百公里急减速次数 | 0.42 | 0.24 | 是 | 是 |
| 每百公里急变速次数 | 2.10 | 2.20 | 否 | 否 |
| 最大连续驾驶时间(h) | 2.26 | 1.97 | 是 | 是 |
| 速度标准差(m/s2) | 20.00 | 22.89 | 是 | 是 |
| 非高速公路行驶的平均速度(m/s) | 31.01 | 32.88 | 否 | 否 |
| 高速公路行驶的平均速度(m/s) | 91.26 | 91.37 | 否 | 否 |

根据样本数据的对比结果，我们选出夜间行驶时间（h）、每百公里急减速次数、行驶总里程（km）、最大连续驾驶时间（h）、速度标准差（m/s2）、行驶总时间（h）六个指标。但由于行驶总里程与行驶总时间高度相关，仅选择其中一个指标即可。行驶总里程最能反映用车强度，因此将行驶总时间这一指标去除。

下面对这5个指标进行熵权分析赋权。

（w1，w2，w3，w4，w5）=（0.1848，0.4445,0.2273,0.1293，0.0141）

由于w5值较小，剔除速度标准差这一指标，重新计算得出：

（w1，w2，w3，w4）=（0.1874，0.4509, 0.2306，0.1312）

## 5.5 建立驾车行为评分模型，对驾车行为进行打分

将100分作为满分，夜间行驶时间（h）、每百公里急减速次数、总里程（km）、最大连续驾驶时间（h）、行驶总时间（h）这四个指标的最大得分分别为19、45、23、13。

设计驾驶行为评分表如表5.4所示：

表5.4 驾驶行为评分表

|  |  |  |  |
| --- | --- | --- | --- |
| **评价指标** | **分值** | **备选项** | **分值** |
| 夜间行驶时间 | 19 | 0.5h以下 | 19 |
| 0.5h（含）~1h | 17 |
| 1h（含）~2h | 15 |
| 2h（含）~4h | 13 |
| 4h（含）~6h | 11 |
| 6h（含）~8h | 9 |
| 8h以上 | 7 |
| 每百公里急减速次数 | 45 | 0 | 45 |
| 1（含）~2 | 40 |
| 2（含）~3 | 35 |
| 3（含）~4 | 30 |
| 4（含）~5 | 25 |
| 5（含）~6 | 20 |
| 6次以上 | 15 |
| 行驶总里程 | 23 | 50km（含）~100km | 23 |
| 100km（含）~300km | 21 |
| 300km（含）~600km | 19 |
| 600km（含）~1000km | 17 |
| 1000km（含）~1500km | 15 |
| 1500km~2000km | 13 |
| 2000km~2500km | 11 |
| 2500km~3000km | 9 |
| 3000km~4000km | 7 |
| 4000km~5000km | 5 |
| 5000km以上 | 3 |
| 最大连续驾驶时间 | 13 | 0.5h以下 | 13 |
| 0.5h~1h | 12 |
| 1h~2h | 11 |
| 2h~4h | 10 |
| 4h~6h | 9 |
| 6h~8h | 8 |
| 8h~10h | 7 |
| 10h~12h | 6 |
| 12h~14h | 5 |
| 14h以上 | 4 |

## 5.6 检验评分模型的有效性

用驾驶行为评分模型计算样本数据的得分情况如表5.5所示：

表5.5 各样本订单驾驶评分

|  |  |  |
| --- | --- | --- |
| 订单ID | 驾驶评分 | 是否出险 |
| 020007100748844 | 83 | 1 |
| 020007360338180 | 91 | 1 |
| 020014799911322 | 94 | 1 |
| 020015826901596 | 93 | 1 |
| 020021181921802 | 92 | 1 |
| 120007211379056 | 89 | 1 |
| 120009349525925 | 78 | 1 |
| 1.2001E+14 | 66 | 1 |
| 120016717906711 | 90 | 1 |
| 220015845101384 | 79 | 1 |
| 320006252425554 | 72 | 1 |
| 320008314131017 | 95 | 1 |
| 320009882501106 | 97 | 1 |
| 320011124931143 | 67 | 1 |
| 320016429691454 | 89 | 1 |
| 320016617541376 | 92 | 1 |
| 320017253711121 | 76 | 1 |
| 320021795801843 | 73 | 1 |
| 420011610421216 | 80 | 1 |
| 520009918291132 | 85 | 1 |
| 520012532971152 | 84 | 1 |
| 520017173551671 | 90 | 1 |
| 520018557621479 | 88 | 1 |
| 620006758278575 | 80 | 1 |
| 720008177639049 | 94 | 1 |
| 720008862029925 | 71 | 1 |
| 720012584601317 | 97 | 1 |
| 720022930411788 | 84 | 1 |
| 920009352071093 | 83 | 1 |
| 920009404761079 | 88 | 1 |
| 920015859781600 | 91 | 1 |
| 920018241591093 | 77 | 1 |
| 920022351261886 | 95 | 1 |
| 020017027411678 | 83 | 0 |
| 020020048901808 | 89 | 0 |
| 020020894361406 | 93 | 0 |
| 020021225741822 | 84 | 0 |
| 020021894981668 | 71 | 0 |
| 120020980841778 | 94 | 0 |
| 220009840938164 | 95 | 0 |
| 220011294119699 | 95 | 0 |
| 220019402951641 | 99 | 0 |
| 220020289041087 | 90 | 0 |
| 220021350271832 | 76 | 0 |
| 320009563706203 | 92 | 0 |
| 320012590439574 | 96 | 0 |
| 320014887741413 | 80 | 0 |
| 420009715321136 | 84 | 0 |
| 420014371331474 | 79 | 0 |
| 420021697471868 | 92 | 0 |
| 520010453755616 | 97 | 0 |
| 520013947071188 | 100 | 0 |
| 520021421941840 | 66 | 0 |
| 520022140881608 | 97 | 0 |
| 620010210365383 | 92 | 0 |
| 620011806501283 | 77 | 0 |
| 620012059966788 | 93 | 0 |
| 620012090035619 | 99 | 0 |
| 720009113715913 | 91 | 0 |
| 720010149248913 | 85 | 0 |
| 720012474371279 | 100 | 0 |
| 720018941291685 | 91 | 0 |
| 720019581901690 | 87 | 0 |
| 820011566066987 | 95 | 0 |
| 820012245391370 | 97 | 0 |
| 820014725457800 | 90 | 0 |
| 820017216741357 | 89 | 0 |
| 820020500451776 | 91 | 0 |

下面对这个评分模型进行效果的评价：

首先，按照驾驶评分的大小对样本进行分类，具体分类标准如下表所示。同时，计算出每组的频数、驾驶评分均值、未出险样本个数、出险样本个数、以及出险样本个数占组频数的比例。

表5.6 驾驶行为评分模型的有效性检测

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **驾驶评分分组** | **频数** | **驾驶评分均值** | **未出险样本个数** | **出险样本个数** | **出险样本个数/组频数** |
| 90（含）~100（含） | 35 | 94.00 | 30 | 5 | 0.142857143 |
| 80（含）~90 | 19 | 84.95 | 13 | 6 | 0.315789474 |
| 70（含）~80 | 11 | 75.36 | 4 | 7 | 0.636363636 |
| 60（含）~70 | 3 | 66.33 | 1 | 2 | 0.666666667 |

由于每组频数的不同，不能仅通过未出险样本数（或出险样本数）来评价该组出险风险的高低，需要用归一化指标（出险样本个数/组频数）来评价该驾驶评分分组的风险大小。

从表中我们可以初步看出，驾驶评分均值越高，该组的出险概率（以出险样本个数/组频数指标来评价）越低。

# 第六章 结论

本文拟定了P2P商业车险定价的基本方法和步骤，同时提供了用户风险甄别的基本方法。

静态从人因素部分本文采用了服从Tweedie分布的广义线性模型对出险金额进行拟合，分别对地区、性别、年龄、驾龄四个因素进行用户风险大小的衡量，发现这四个因素对被保险人风险大小的衡量均有显著作用。地区方面，北京出险风险大于上海；性别方面，男性出险风险远大于女性；从年龄来看，21~24岁的被保险人风险最高，40岁以上的被保险人风险其次，而25~29岁年龄段的风险相对较低；从驾龄来看，驾龄小于1年的被保险人风险最大，其次是驾龄在4~7年的被保险人，而7年以上驾龄的被保险人，其风险会随着驾龄的增加而降低。

动态从人因素部分，利用P2P车险平台天然的数据优势，本文对用户驾车动态行为进行分析，采用熵权法对用户驾车行为进行打分。其中，每百公里急减速次数这一行为因子指标对出险风险的影响最大，夜间行驶时间、行驶总里程以及最大连续驾驶时间这三个行为因子对风险的高低也有比较显著的影响。

这两部分的模型最终的拟合度都比较好，说明用此方法对租客（被保险人）的风险判别是有效的。然而，进一步根据用户风险的大小与具体车险费率挂钩就需要进一步综合保险公司的理赔成本和运营成本等成本来进行费率的厘定，本文不再对此进行讨论研究。

同时，针对此模型的正向激励作用未来也有待于进一步研究。

# 参考文献

1. 董志勇.行为金融学[M].北京:北京大学出版社.2009.
2. 李心丹.行为金融理论:研究体系及展望[J].金融研究,2005,(3).
3. 段白鸽,余东发,张连增.国外车险里程定价理论与实践[J].保险研究,2012, (2).
4. 段冀阳,李志忠,驾驶行为模型的研究进展[J].中国安全科学学报,2012,22(2).
5. 郑东鹏,蒋祖华,章倩.驾驶员风险驾驶行为及影响因素分析[J]. 人类功效学, 2014, 20(1).
6. 张连增, 段白鸽. 行驶里程数对车险净保费的影响研究——基于公路里程对交通事故损失的影响视角[J]. 保险研究, 2012, (6).
7. 袁小康. 我国车险奖惩系统的实证研究[D]. 西南财经大学,2011.
8. 曾辉, 范兴华. 关于我国机动车辆保险费率厘定因素的思考[J]. 经济师, 2003:159-159..
9. 曾里斌.广义线性模型在非寿险费率厘定中的运用[J].金融经济(理论版), 2007:60-61.
10. 赵慧卿,王汉章.我国车险费率厘定的实证研究——基于广义线性模型的分析[J]. 天津商业大学学报, 2011 : 8-12.
11. 朱曦. 数据挖掘在非寿险分类费率厘定中的运用研究[D]. 湖南大学, 2005.
12. 郁佳敏.汽车保险精算定价模型研究综述[J].上海金融学院学报. 2010:14-19.
13. 张晰旸.按里程付费的车险研究[D].西南财经大学, 2014.
14. 韩成卉,赵绰翔,郑苏晋.“行为定价保险”：车险定价与驾驶行为理论[J].金融发展研究,2015,(8).
15. 孙继伟,张连增.ZAIG模型在车险定价中的应用研究[J].保险研究,2013,(4).
16. 庄明科,白海峰,谢晓非.驾驶人员风险驾驶行为分析及相关因素研究[J].北京大学学报（自然科学版）,2008,44(3).
17. 朱爽.车联网环境下基于UBI的车险费率厘定模式与方法研究[D]. 北京交通大学, 2015.
18. 张茵.对加入从人因素的我国车险费率厘定的研究[D],湖南大学,2007.
19. 曾娟.机动车辆保险分类费率厘定原理与方法研究[D],湖南大学,2007.
20. 孙维伟.基于Tweedie类分布的广义可加模型在车险费率厘定中的应用[J],天津商业大学学报,2014,(1).
21. 吴开亚,金菊良. 区域生态安全评价的熵组合权重属性识别模式[J],地理科学,2008,(12).
22. 张献国,张学军.基于熵权系数法的信息系统安全模糊风险评估[J],内蒙古大学学报,2005,(11).
23. 杨达.考虑后车的车辆跟驰行为建模及分析[D],西南交通大学,2013
24. 刘彦宇.应用心理学[D],浙江理工大学,2013
25. 杨正.基于大数据架构的智能交通解决方案[J].北京联合大学学报(自然科学版),2014,28(4).
26. 杨伯卿.中国移动在车联网运营中发展策略探析[J]. [软件](http://c.wanfangdata.com.cn/Periodical-ranj.aspx),2013,34(5).
27. 徐永兵.解读“车联网”[J].商用汽车,2012,(16).
28. 孟生旺,张连增.刘乐平.精算学基础[M].北京:中国人民大学出版社,2016:99-135.
29. 孟生旺,刘乐平.肖争艳.非寿险精算学[M].北京:中国人民大学出版社,2015:196-254.
30. 王晓军,孟生旺.保险精算学[M].北京:中国人民大学出版社,2006:256-278.
31. 郑苏晋.车联网与车险风险管理[M].北京:经济科学出版社,2015:116-124.
32. 沈树茂.户外网桥远程维护与监控系统的开发[D].苏州:苏州大学,2010.
33. Anderson D, Feldblum S, Modlin C, et al. A practitioner’s guide to generalized linear models[J].Casualty Actuarial Society Discussion Paper Program, 2004:1-116.
34. Butler P. Automobile insurance pricing: operating cost versus ownership cost; the implications for women[C].Proceedings, Women’s Travel Issues Second National Conference, 1996:737-51.
35. Clark D R, Thayer C A. A primer on the exponential family of distributions [C].Casualty Actuarial Society Spring Forum, 2004:117-148.
36. Finger R J.Risk classification[J].Foundations of Casualty Actuarial Science, 2006: 231-276.
37. Lereah D A. Insurance markets: information problems and regulation [M]. Praeger,1985.
38. Guensler, Randall, Amekudzi, Adjo, Jennifer Williams, Shannon Mergelsberg, Jennifer Ogle. Current State Regulatory Support for Pay-as-You-Drive Automobile Insurance Options［J］.Journal of Insurance Regulation, NAIC，2003,21( 3).
39. Khazzoom, J.Daniel. Pay-at-the-Pump Auto Insurance: Review of Criticisms and Proposed Modification［J］. Journal of Insurance Regulation，2000，18( 4) : 448 -496．
40. Pigou, A.C. The Economics of Welfare 1952［M］.4th edition，Macmillan，London，1920.
41. Sugarman，Stephen D. Face It: The Current System for Injury Liability is a Mess［R］. The Sacramento Bee, 1990,April 1．
42. Tweedie M C K. An Index Which Distinguishes Between Some Important Exponential Families [C]//Ghosh J K, Roy J. Statistics:Applicetions and New Direction. Proceeding of the Indian Statistical Institute Golden JubileeInternational Conference,1984:579-604.
43. De Jong P, Heller G Z. Generalized Linear Models for Insurance Data[M].Cambrige:Cambrige University Press,2008.
44. Haste T,Tibshirani R.Generalized Additive Models[M].London:Chapman and Hall,1990.
45. Agresti,Alan.Categorical Data Analysis[M].New York:John Wiley,1990.
46. Ludwig Fahrmeir Gerhard Tutz.Multivariate Statistical Modelling Based on Generalized Linear Models[M].Springer-Verlag New York,Inc.1994.
47. Green.W.H.Econometric Analysis[M].New York:Macmillan.1990.
48. Hadaller D,Keshav S, Brecht T,et al. Vehicular Opportunistic Communication Under the Microscope [C]//Proceeding of the 5th International Conference on Mobile System, Applications and Service.USA:[s.n.],2007,USA.
49. Akyildiz I F, Wang Xudong,Wang Weilin.Wireless meshnetworks:a survey[J].Computer Networks,2005,47(4):445-487.
50. Brown B.,Chui M. and Manyika J..Are You Ready for the Era of “Big Data”?[J].McKinsey Quarterly,2011,(4).

# 附录

## A 重要代码

A.1 Python处理动态数据部分的代码

import os

from math import \*

import re

from math import radians, cos, sin, asin, sqrt

import time

def scan\_files(directory,prefix=None,postfix=None):

files\_list=[]

for root, sub\_dirs, files in os.walk(directory):

for special\_file in files:

if postfix:

if special\_file.endswith(postfix):

files\_list.append(os.path.join(root,special\_file))

elif prefix:

if special\_file.startswith(prefix):

files\_list.append(os.path.join(root,special\_file))

else:

files\_list.append(os.path.join(root,special\_file))

return files\_list

# input Lat\_A 纬度A

# input Lng\_A 经度A

# input Lat\_B 纬度B

# input Lng\_B 经度B

# output distance 距离(km)

def calcDistance(Lat\_A, Lng\_A, Lat\_B, Lng\_B):

ra = 6378.140 # 赤道半径 (km)

rb = 6356.755 # 极半径 (km)

flatten = (ra - rb) / ra # 地球扁率

rad\_lat\_A = radians(Lat\_A)

rad\_lng\_A = radians(Lng\_A)

rad\_lat\_B = radians(Lat\_B)

rad\_lng\_B = radians(Lng\_B)

pA = atan(rb / ra \* tan(rad\_lat\_A))

pB = atan(rb / ra \* tan(rad\_lat\_B))

xx = acos(sin(pA) \* sin(pB) + cos(pA) \* cos(pB) \* cos(rad\_lng\_A - rad\_lng\_B))

c1 = (sin(xx) - xx) \* (sin(pA) + sin(pB)) \*\* 2 / cos(xx / 2) \*\* 2

c2 = (sin(xx) + xx) \* (sin(pA) - sin(pB)) \*\* 2 / sin(xx / 2) \*\* 2

dr = flatten / 8 \* (c1 - c2)

distance = ra \* (xx + dr)

return distance

# Lat\_A=32.060255; Lng\_A=118.796877 # 南京

# Lat\_B=39.904211; Lng\_B=116.407395 # 北京

# distance=calcDistance(Lat\_A,Lng\_A,Lat\_B,Lng\_B)

# print('(Lat\_A, Lng\_A)=({0:10.3f},{1:10.3f})'.format(Lat\_A,Lng\_A))

# print('(Lat\_B, Lng\_B)=({0:10.3f},{1:10.3f})'.format(Lat\_B,Lng\_B))

# print('Distance={0:10.3f} km'.format(distance))

def haversine(lon1, lat1, lon2, lat2): # 经度1，纬度1，经度2，纬度2 （十进制度数）

"""

Calculate the great circle distance between two points

on the earth (specified in decimal degrees)

"""

# 将十进制度数转化为弧度

lon1, lat1, lon2, lat2 = map(radians, [lon1, lat1, lon2, lat2])

# haversine公式

dlon = lon2 - lon1

dlat = lat2 - lat1

a = sin(dlat/2)\*\*2 + cos(lat1) \* cos(lat2) \* sin(dlon/2)\*\*2

c = 2 \* asin(sqrt(a))

r = 6371 # 地球平均半径，单位为公里

return c \* r \* 1000

# print(haversine(Lng\_A,Lat\_A,Lng\_B,Lat\_B))

def Checktime\_atNight(ctime):

Flag=0

#将ctime转化为时间戳

ts=time.mktime(time.strptime(ctime,'%Y-%m-%d %H:%M:%S'))

#将时间戳转化为localtime

ct = time.localtime(ts)

hour = time.strftime('%H',ct)

if int(hour)<= 19 and int(hour)>=6:

Flag=0

else:

Flag=1

return Flag

# print(Checktime\_atNight('2015-08-29 06:45:25'))

import os

from math import \*

import pandas as pd

import numpy as np

import re

import time

import car\_project\_lib as car

directory = 'C:\项目文件\GPS轨迹数据\\' # 'C:\项目文件\GPS轨迹数据\_未出险车辆\_对比样本\\'

directory\_data = 'C:\项目文件\GPS轨迹数据\_已处理\\' # 'C:\项目文件\GPS轨迹数据\_未出险车辆\_对比样本\_已处理\\'

directory\_stat = 'C:\项目文件\\'

file\_list = car.scan\_files(directory, postfix='.xlsx') # 读取原始数据

data\_stat = pd.DataFrame() # 记录统计数据

for f\_name in file\_list:

print(f\_name)

ID = re.findall(r"\d+\.?\d\*", f\_name)[0]

df = pd.read\_excel(f\_name, encoding='gbk')

size = df['速度'].size

# calculate the distance\_delta

distance\_delta = pd.Series(np.zeros(size), name='distance\_delta')

for i in range(size):

if i == 0:

distance\_delta[i] = 0

else:

distance\_delta[i] = car.haversine(df['经度'][i - 1], df['纬度'][i - 1], df['经度'][i], df['纬度'][i])

df['distance\_delta'] = distance\_delta

# time\_delta

time\_delta = pd.Series(np.zeros(size), name='time\_delta')

for i in range(size):

if i == 0:

time\_delta[i] = 0

else:

time\_delta[i] = time.mktime(time.strptime(str(df['时间'][i]), '%Y-%m-%d %H:%M:%S')) \

- time.mktime(time.strptime(str(df['时间'][i - 1]), '%Y-%m-%d %H:%M:%S'))

df['time\_delta'] = time\_delta

# at\_night

at\_night = pd.Series(np.zeros(size), name='at\_night')

for i in range(size):

at\_night[i] = car.Checktime\_atNight(str(df['时间'][i]))

df['at\_night'] = at\_night

# calculate continuous time

continuous\_time = pd.Series(np.zeros(size), name='continuous\_time')

for i in range(size):

if i == 0:

continuous\_time[i] = 0

else:

if (df['速度'][i] == 0) and (df['速度'][i - 1] == 0):

continuous\_time[i] = 0

else:

delta\_time = time.mktime(time.strptime(str(df['时间'][i]), ' %Y-%m-%d %H:%M:%S')) \

- time.mktime(time.strptime(str(df['时间'][i - 1]), '%Y-%m-%d %H:%M:%S'))

continuous\_time[i] = continuous\_time[i - 1] + delta\_time

df['continuous\_time'] = continuous\_time

# calculate acceleration

acceleration = pd.Series(np.zeros(size), name='acceleration')

for i in range(size):

if i == 0:

acceleration[i] = 0

else:

acceleration[i] = (df['速度'][i] - df['速度'][i - 1]) \* 1000 / 3600 / time\_delta[i]

df['acceleration'] = acceleration

# 判断高速公路行驶与非高速公路行驶。

# 判断依据：连续5分钟超过70KM/小时，则判定为高速公路驾驶

highWay = pd.Series(np.zeros(size), name='highWay')

for i in range(size):

time\_used = time.mktime(time.strptime(str(df['时间'][i]), '%Y-%m-%d %H:%M:%S')) \

- time.mktime(time.strptime(str(df['时间'][0]), '%Y-%m-%d %H:%M:%S'))

if time\_used <= 300:

continue

for j in range(i, -1, -1):

time\_used\_5min = time.mktime(time.strptime(str(df['时间'][i]), '%Y-%m-%d %H:%M:%S')) \

- time.mktime(time.strptime(str(df['时间'][j]), '%Y-%m-%d %H:%M:%S'))

if time\_used\_5min > 300:

# 判断[j]~[i]这段时间内（300秒），是否一直以大于70KM的时速行驶

flag = True

for k in range(j, i + 1, 1):

if df['速度'][k] < 70:

flag = False

# 确认是在高速公路

if flag:

for m in range(j, i + 1, 1):

highWay[m] = 1

break

df['highWay'] = highWay

df.to\_excel(directory\_data + ID + '.xlsx', index=False) # 保存已经处理过的数据

# --------------------------------------------------------------------------------------

# --------------------------------------------------------------------------------------

# 速度大于零的数据

df\_select = df[df['速度'] != 0]

size\_select = df\_select['速度'].size

# 夜间行驶的数据

df\_select\_atNight = df\_select[df\_select['at\_night'] == 1]

# 高速公路行驶的数据

df\_select\_highWay = df\_select[df\_select['highWay'] == 1]

df\_select\_highWay.index = np.arange(df\_select\_highWay['速度'].size)

# 非高速公路行驶的数据

df\_select\_notHighWay = df\_select[df\_select['highWay'] == 0]

df\_select\_notHighWay.index = np.arange(df\_select\_notHighWay['速度'].size)

# 行驶总时间；计量单位：天

time\_total = (time.mktime(time.strptime(str(df['时间'][size - 1]), '%Y-%m-%d %H:%M:%S')) - time.mktime(

time.strptime(str(df['时间'][0]), '%Y-%m-%d %H:%M:%S'))) / 3600 / 24

# 总里程数，单位：KM

s\_total = np.sum(df['distance\_delta']) / 1000

# 计算急加速与急减速的次数，判断标准为：正加速度大于1.38米每秒为急加速；

# 负加速度小于-1.5米每秒为急减速。

# 注意：这里计算频次依赖于加速度计的采样频率。后面再把频次数据用每100KM的里程来归一化。

num\_rapid\_acceleration = 0

num\_rapid\_slowDown = 0

for i in range(size):

if df['acceleration'][i] > 1.38:

num\_rapid\_acceleration += 1

if df['acceleration'][i] < -1.5:

num\_rapid\_slowDown += 1

# 计算高速公路/非高速公路行驶过程中，超速的次数。

# 判断标准为：高速公路时速大于120KM；非高速公路时速大于70KM。

# 注意：这里计算次数依赖于速度传感器的采样频率。后面再把频次数据用每100KM的里程来归一化。

num\_Overspeed\_highWay = 0

num\_Overspeed\_notHighWay = 0

for i in range(df\_select\_highWay['速度'].size):

if df\_select\_highWay['速度'][i] > 120:

num\_Overspeed\_highWay += 1

for i in range(df\_select\_notHighWay['速度'].size):

if df\_select\_notHighWay['速度'][i] > 70:

num\_Overspeed\_notHighWay += 1

# 高速公路行驶的平均速度

v\_average\_highWay = nan

if df\_select\_highWay['速度'].size>0:

v\_average\_highWay = np.sum(df\_select\_highWay['速度']) / df\_select\_highWay['速度'].size

# 非高速公路行驶的平均速度

v\_average\_notHighWay = nan

if df\_select\_notHighWay['速度'].size>0:

v\_average\_notHighWay = np.sum(df\_select\_notHighWay['速度']) / df\_select\_notHighWay['速度'].size

data\_id = {'ID': pd.Series([ID]),

'租车总时间【天】': pd.Series([time\_total]),

'总里程【公里】': pd.Series([s\_total]),

'行驶总时间【小时】': pd.Series([np.sum(df\_select['time\_delta']) / 3600]),

'夜间行驶总时间【小时】': pd.Series([np.sum(df\_select\_atNight['time\_delta']) / 3600]),

'单日行驶里程【公里】': pd.Series([np.sum(df['distance\_delta']) / 1000 / time\_total]),

'单日行驶时间【小时】': pd.Series([np.sum(df\_select['time\_delta']) / 3600 / time\_total]),

'单日夜间行驶时间【小时】': pd.Series([np.sum(df\_select\_atNight['time\_delta']) / 3600 / time\_total]),

'平均行驶速度【公里每小时】': pd.Series([np.sum(df\_select['速度']) / size\_select]),

'最高速度【公里每小时】': pd.Series([max(df\_select['速度'])]),

'平均加速度【米每秒2】': pd.Series([np.sum(abs(df\_select['acceleration'])) / size\_select]),

'最大加速度【米每秒2】': pd.Series([max(abs(df\_select['acceleration']))]),

'最大连续驾驶时间【小时】': pd.Series([max(df['continuous\_time']) / 3600]),

'速度标准差': pd.Series([np.std(df\_select['速度'])]),

'急加速次数【每100KM】': pd.Series([num\_rapid\_acceleration / s\_total \* 100]),

'急减速次数【每100KM】': pd.Series([num\_rapid\_slowDown / s\_total \* 100]),

'急速变速次数【每100KM】': pd.Series([(num\_rapid\_slowDown + num\_rapid\_acceleration) / s\_total \* 100]),

'高速公路行驶中超速次数【每100KM】': pd.Series(num\_Overspeed\_highWay / s\_total \* 100),

'非高速公路行驶中超速次数【每100KM】': pd.Series(num\_Overspeed\_notHighWay / s\_total \* 100),

'高速公路行驶的平均速度': pd.Series([v\_average\_highWay]),

'非高速公路行驶的平均速度': pd.Series([v\_average\_notHighWay])

}

df\_id = pd.DataFrame(data\_id)

data\_stat = data\_stat.append(df\_id, ignore\_index=True)

data\_stat.to\_excel(directory\_stat + '统计数据.xlsx', index=False)

A.2 Matlab进行熵权法赋权部分的代码

function weights = EntropyWeight(R)  
%% 熵权法求指标权重,R为输入矩阵,返回权重向量weights  
  
[rows,cols]=size(R);   % 输入矩阵的大小,rows为对象个数，cols为指标个数  
k=1/log(rows);         % 求k  
  
f=zeros(rows,cols);    % 初始化fij  
sumBycols=sum(R,1);    % 输入矩阵的每一列之和(结果为一个1\*cols的行向量)  
% 计算fij  
for i=1:rows  
    for j=1:cols  
        f(i,j)=R(i,j)./sumBycols(1,j);  
    end  
end  
  
lnfij=zeros(rows,cols);  % 初始化lnfij  
% 计算lnfij  
for i=1:rows  
    for j=1:cols  
        if f(i,j)==0  
            lnfij(i,j)=0;  
        else  
            lnfij(i,j)=log(f(i,j));  
        end  
    end  
end  
  
Hj=-k\*(sum(f.\*lnfij,1)); % 计算熵值Hj  
weights=(1-Hj)/(cols-sum(Hj));  
end

## B 数据处理&分析工具

1. SPSS官网

http://www-01.ibm.com/software/cn/analytics/spss/

1. MATLAB官网

http://cn.mathworks.com/?requestedDomain=www.mathworks.com

1. PYTHON官网

https://www.PYTHON.org/

# 致谢

本文论文的选题是我在实习过程中产生的，目前随着我国车险费率厘定市场化进程的不断深入以及车联网设备的逐步普及，我觉得建立一套利用驾驶动态数据来判断被保险人风险大小的模型具有非常重要的实践意义，于是我选择了这个方向进行论文创作。

首先要感谢我的导师李杰教授对我此次论文的帮助，此论文能够顺利完成得益于李老师的辛勤付出。在选题过程中李老师对我的研究方向的给予了很大的支持，同时在论文写作过程中时常在思路上给我引导和启发。李老师在研究方面思路清晰、洞察力敏锐，时常帮我拨开迷雾，让我能够一直走向正确的研究方向。李老师严谨求实的治学态度、精益求精的工作作风、正直诚恳、与人为善的待人风范对我影响极其深远，将指引我今后的工作与生活。

同时，还要感谢宝驾租车CEO李如彬先生以及宝驾租车的所有同事，他们全力协助我此次论文的研究工作，为我提供了大量的项目研究数据，使我在项目实践中能够顺利开展。

我也要向与我朝夕相处三年的同学们表示感谢。在学习上我们互相探讨，解决课程或科研问题，在生活中，互相帮助，共同成长。其中，要特别感谢许浩同学在我此次论文的写作过程中给予我许多思路上的启发与技术上的支持。

校园生活总是最美好的，研究生生涯一晃而过，开学那天还犹如昨天，今天的我们要即将步入社会。在学校的这三年中，不仅在学习上学到了很多，在眼界、胸怀上都感觉变得宽阔了。因为遇到了各位优秀的老师，和一群有理想有志向的同学，我的研究生生活变得丰富充实，我从他们身上也学习到了很多东西，使我受益匪浅，我相信，我会记得各位老师对我们的谆谆教导，记得我们许诺共同努力的目标，希望能够在未来的社会工作中继续认真努力的脚步，踏踏实实的为社会贡献我的一份力量，争取为母校增添一份光彩，为社会做出一份贡献。

# 北京大学学位论文原创性声明和使用授权说明

**原创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

论文作者签名： 日期： 年 月 日

**学位论文使用授权说明**

（必须装订在提交学校图书馆的印刷本）

本人完全了解北京大学关于收集、保存、使用学位论文的规定，即：

* 按照学校要求提交学位论文的印刷本和电子版本；
* 学校有权保存学位论文的印刷本和电子版，并提供目录检索与阅览服务，在校园网上提供服务；
* 学校可以采用影印、缩印、数字化或其它复制手段保存论文；
* 因某种特殊原因需要延迟发布学位论文电子版，授权学校□一年/□两年/□三年以后，在校园网上全文发布。

（保密论文在解密后遵守此规定）

论文作者签名： 导师签名：

日期： 年 月 日