

本科毕业设计（论文）

**基于大数据的地铁新技术采用与居民出行选择关系研究**

**Research on the relationship between new metro technology adoption and residents’ travel choice based on big data**

学 院： 经管学院

专 业： 信息管理与信息系统

学生姓名： 张露

学 号： 18711076

指导教师： 宫大庆

**北京交通大学**

2022年5月

学士论文版权使用授权书

本学士论文作者完全了解北京交通大学有关保留、使用学士论文的规定。特授权北京交通大学可以将学士论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名： 指导教师签名：55c4e57002d1b8cc72192489de361ee



签字日期： 2022 年 5 月 30 日 签字日期： 2022 年 5 月 30 日

中文摘要

随着我国总人口数目的不断增加，国民经济的不断增长，地铁作为一种绿色环保、便捷高效的公共交通方式已越来越成为居民出行的主流选择。随着“互联网+”移动技术的发展和现代社会生活节奏的加快，《关于积极推进“互联网+”行动的指导意见》中明确指出轨道交通行业势必要经历“互联网+”的科技变革。通过应用“互联网+”等新技术，轨道交通行业使居民出行过程变得更加便捷且环保。同时，在我国发改委颁布的《关于促进绿色消费的指导意见》中指出，要积极引领居民实施健康绿色的生活方式和消费模式。综上所述，地铁新技术的使用和绿色出行的实践都是人类发展的必然趋势。因此如果能将二者联系起来，不仅可以为地铁新技术的未来发展作指导，还能够对绿色环保的生活方式及消费模式产生积极的影响。

为探寻地铁新技术与绿色出行之间的关系，我们需要考虑居民出行方式的选择在这之间产生的重要作用。因此本文的研究问题是探究地铁新技术（比如二维码刷码过闸）的使用是否能对居民出行方式的选择产生影响，从而促使居民选择更加绿色的公共交通工具（比如地铁）作为出行方式。本文主要以二维码刷码过闸机的使用为例，采用实证研究的方法，通过对政策颁布前后大量居民出行的数据进行统计分析，时空模式挖掘，Logistic回归模型和结构方程模型的建立以及控制变量来探究地铁新技术的使用对居民出行选择是否产生影响，从而得出地铁新技术的使用是否会促进更多居民选择绿色出行这一结论。

本文通过研究发现，在地铁新技术的相关政策颁布以后，居民的客流量以及选择乘坐地铁出行的意愿程度有所增加，因此能够得出地铁新技术的使用会对居民选择乘坐地铁出行产生积极的影响，这将促使居民选择更加绿色的公共交通工具（比如地铁）作为出行方式这一结论。本文旨在为绿色出行和科技变革在轨道交通行业中的推进提供一些普适性的建议，并且也对传统居民出行行为进行了扩展，不但弥补了居民出行行为在地铁行业的科研缺口，而且还将未来科技出行领域与绿色环保领域之间构建起了密切联系。

**关键词：**地铁新技术；出行方式选择；时空模式挖掘；logistic回归；绿色出行

ABSTRACT

With the continuous increase of China’s total population and the continuous growth of national economy, subway, as a green, convenient and efficient public transportation mode, has become the mainstream choice for residents to travel. With the development of “Internet +” mobile technology and the acceleration of the pace of modern social life, it is clearly pointed out in the “Guidance on Actively Promoting Internet + Action” that the rail transit industry is bound to experience “Internet +” technological change. By applying new technologies such as “Internet Plus”, the rail transit industry has made the journey for residents more convenient and environmentally friendly. At the same time, the Guidance on Promoting green consumption issued by China’s National Development and Reform Commission points out that it is necessary to actively lead residents to implement healthy and green lifestyle and consumption patterns. To sum up, the use of new subway technology and the practice of green travel are inevitable trends of human development. Therefore, if the two can be connected, it can not only guide the future development of subway new technology, but also have a positive impact on the green lifestyle and consumption pattern.

In order to explore the relationship between new subway technology and green travel, we need to consider the important role of residents’ choice of travel mode in this relationship. Therefore, the research question of this paper is to explore whether the use of new subway technology (such as QR code swipe code to pass the gate) can have an impact on residents’ choice of travel mode, so as to encourage residents to choose greener public transport (such as subway) as their travel mode. This article mainly by QR code brush code lock machine as an example, the use of the empirical research method, before and after the policy issued by a large number of residents travel to statistical analysis of data, spatial and temporal pattern mining, the establishment of the logistic regression model and structural equation model to explore the new technology and control variables affect residents’ travel choice whether or not. It is concluded that whether the use of new subway technology will promote more residents to choose green travel.

Through the research, this paper found that after the promulgation of the relevant policies of subway new technology, the passenger flow of residents and their willingness to choose subway travel increased. Therefore, it can be concluded that the use of subway new technology will have a positive impact on residents’ choice of subway travel. This will encourage residents to choose greener public transport (such as the subway) as a means of travel. This paper aims at providing a green travel and technological change in the rail transit industry to promote provide some general advice, and also extends to traditional residents travel behavior, not only makes up for the residents’ travel behavior in metro industry research gap, but also will be the future travel areas of science and technology and to build up a close relationship between green fields.

**KEYWORDS：**New subway technology; Choice of travel mode; Spatio-temporal pattern mining; Logistic regression; Green travel

目 录

[中文摘要 i](#_Toc102588762)

[ABSTRACT ii](#_Toc102588763)

[目 录 iv](#_Toc102588764)

[1 引言 1](#_Toc102588765)

[1.1 研究背景与研究问题 1](#_Toc102588766)

[1.1.1 研究背景 1](#_Toc102588767)

[1.1.2 研究问题 2](#_Toc102588768)

[1.2 研究方法 2](#_Toc102588769)

[1.3 研究内容与步骤 3](#_Toc102588770)

[1.4 研究意义与价值 5](#_Toc102588771)

[1.4.1 理论意义 5](#_Toc102588772)

[1.4.2 实践价值 5](#_Toc102588773)

[2 理论基础与文献综述 6](#_Toc102588774)

[2.1 理论基础 6](#_Toc102588775)

[2.1.1 地铁新技术界定 6](#_Toc102588776)

[2.1.2 地铁出行方式特性分析 8](#_Toc102588777)

[2.1.3 聚类分析 10](#_Toc102588778)

[2.2 国内外研究现状 12](#_Toc102588779)

[2.2.1 出行方式选择影响因素 12](#_Toc102588780)

[2.2.2 公共交通新技术对居民出行的影响 18](#_Toc102588781)

[2.2.3 模型在出行选择方面的应用 20](#_Toc102588782)

[2.3 文献述评 24](#_Toc102588783)

[3基于大数据的地铁乘客行为模式挖掘 25](#_Toc102588784)

[3.1 数据处理 25](#_Toc102588785)

[3.1.1 数据获取及说明 25](#_Toc102588786)

[3.1.2 基于Python的数据整合 26](#_Toc102588787)

[3.2 基于客流量统计的乘客行为分析 27](#_Toc102588788)

[3.3 基于时间差变化的乘客行为分析 32](#_Toc102588789)

[3.3.1 时间差指标及获取 32](#_Toc102588790)

[3.3.2 基于时间差变化的统计分析 34](#_Toc102588791)

[3.4 基于空间模式挖掘的乘客行为分析 39](#_Toc102588792)

[3.4.1 活跃乘客及活跃天数计算 39](#_Toc102588793)

[3.4.2 基于活跃乘客及天数的聚类分析 41](#_Toc102588794)

[3.5 研究结果分析 44](#_Toc102588795)

[4居民出行方式选择行为模型构建 46](#_Toc102588796)

[4.1 居民出行调查 46](#_Toc102588797)

[4.1.1 思路设计 46](#_Toc102588798)

[4.1.2 问卷设计 46](#_Toc102588799)

[4.1.3 居民出行调查问卷数据整理 47](#_Toc102588800)

[4.2 Logistic回归模型分析 49](#_Toc102588801)

[4.2.1 模型的选择 49](#_Toc102588802)

[4.2.2 模型的建立 51](#_Toc102588803)

[4.3 结构方程模型分析 53](#_Toc102588804)

[4.3.1 模型的选择 53](#_Toc102588805)

[4.3.2 模型的建立 55](#_Toc102588806)

[4.4 控制变量分析 57](#_Toc102588807)

[4.4.1 内生变量与解释变量关系 57](#_Toc102588808)

[4.4.2 外生变量与因变量关系 58](#_Toc102588809)

[4.5 研究结果分析 59](#_Toc102588810)

[5相关政策建议 60](#_Toc102588811)

[5.1 其他地铁新技术政策分析 60](#_Toc102588812)

[5.2 相关政策建议 63](#_Toc102588813)

[6研究结论 64](#_Toc102588814)

[参考文献 65](#_Toc102588815)

[致 谢 68](#_Toc102588816)

[附 录 69](#_Toc102588817)

1. 引言

1.1 研究背景与研究问题

* + 1. 研究背景

随着“互联网+”移动通信技术的发展和现代社会生活节奏的加快，轨道交通行业也通过应用互联网技术，将居民出行过程逐渐移植到了移动环境中。2015 年 7 月 1 日国务院办公厅印发的《关于积极推进“互联网+”行动的指导意见》，该文章提出，通过“互联网+”作为技术基础来便利城市交通，并支持网络信息平台为居民出行服务提供即时城市交通运行状态自动查询、出行路线规划、网上自助购票等服务，推动基于互联网平台的出行方式信息化对接服务和一站式服务[1]。此意见提出后，网上云购票、二维码过闸、NFC以及银联等新一代移动支付技术被运用到轨道交通行业中，形成了在支付过闸方面的地铁新技术并得到了迅速推广。

2018年，北京市地铁线全网（不含西郊线）实现了二维码刷码乘坐地铁，标志着北京地铁全线步入了只需手机刷码乘坐地铁的“新时代”。此外，地铁的刷码支付过闸引入了双离线技术和安全加密技术的智能地铁新技术，这使扫码流程不再需要依靠网络环境，平均响应时间不会超过0.2秒，可以真正适应北京市公共交通客流量高峰期的快速地铁通行需求[2]。因此，二维码刷码过闸技术作为地铁新技术，有效地解决了地铁自动售检票存在的车票消耗成本高、购票效率低以及不够环保等问题，同时也丰富了地铁票务的支付方式[2]。除了二维码过闸、空调集中供冷和地铁站台屏蔽门等已经较为成熟的地铁新技术，还有2021年开始投入站点试测的乘客快速进站技术以及在2022年冬奥会中使用的智慧运维、MBAS、MFAS、MSCADA、云平台以及三码合一等建设性地铁新技术的应用。因此轨道交通行业势必要进行并将长期处于“互联网+”的科技改革中。

另外，由于我国经济的迅速发展以及科技变革所带来的便利条件，城市居民的出行量也随之大幅度提升，由此产生的交通碳排放污染给城市居民生活和环境带来了极坏的影响。根据全球能源署(IEA)的统计资料,二零零七年在交通方面所产生的总碳排放量,大约为世界碳排放量总量的百分之二十三,并估计该比率将于二零三零年上升至百分之四十一[3]。因此，交通运输行业的碳排放量问题已成为全球瞩目的焦点议题。所以，降低交通行业的碳排放量变成了绿色环保的首要任务，而居民出行方式的选择将成为决定交通碳排放水平的关键因素。国务院发改委颁布的《关于促进绿色消费的指导意见》也明确提出了积极引导城市居民采用绿色消费模式和生活方式，并提倡徒步、自行车和公共交通等绿色生态的生活模式[4]。而地铁作为一种绿色环保、便捷高效、安全且可以节省大部分地上空间的城市公共交通方式，将日益引起人们关注，越来越成为大型都市中分担公共交通乘客的主要交通方式。

因此，地铁新技术的使用以及绿色出行的推广都是人类发展的必然趋势，而居民出行选择则是联系这两者之间的关键因素。

* + 1. 研究问题

基于上述研究背景，如果将地铁新技术与绿色出行联系起来，能够对绿色环保的生活方式及消费模式产生积极的影响。为探寻这一关系，我们需要考虑居民出行方式的选择在这之间产生的重要作用。因此本文的研究问题是探究地铁新技术（比如二维码刷码过闸）的使用是否能对居民出行方式的选择产生影响，从而促使居民选择更加绿色的公共交通工具（比如地铁）作为出行方式。

* 1. 研究方法

本文采用文献回顾法、实证研究法、数量研究法三种方法，对地铁新技术与居民出行选择的关系进行研究。

（1）文献回顾法。通过在 Proquest、知网、百度学术、Summon 等文献数据库中搜索，阅读了解了大量有关地铁新技术、居民出行选择、时空模式挖掘（聚类分析法）、Logistic回归模型及内生性因素分析等方面的文献。同时，通过对文献进行分类整理，加深了对文献资料的理解和认识。

（2）实证研究法。通过将数据和实验手段相结合从而得出结论，并且该结论在同一条件下具有可证性。本文主要是对二维码刷码过闸与居民出行选择的关系进行实验和调查，从而探究地铁新技术与居民出行选择之间存在的普遍联系，揭示出地铁新技术对居民绿色出行的影响。

（3）数量研究法。该方法又经常被叫做“定量研究法”和“统计研究法”，是通过分别研究测量对象的规模、数量、频率、深度和各种数学关系，认识并明确事物间存在的内在关系、规律以及未来趋势，以实现所预期目的的一类科学研究方法。本文将基于调查法来调查影响居民出行选择的因素，然后利用Logistic模型、控制变量等多元的统计方法进行分析研究，进而获得地铁新技术对居民出行选择的重要影响

* 1. 研究内容与步骤

本文属于实证型研究，利用Python和SPSS技术对大量月度地铁数据进行数量分析、时空模式挖掘（聚类分析）、模型建立、控制变量以及稳健性因素分析，来挖掘出地铁新技术的使用对居民出行选择带来的影响，从而对未来的绿色出行发展及轨道交通行业的科技发展方向提供参考。本文的研究内容及研究步骤如下：

第一部分，引言。本部分从研究背景引入，分析研究问题和解决问题的意义所在，再通过对研究方法的介绍，梳理出本文的研究步骤和技术路线。

第二部分，理论基础与文献综述。理论基础部分将从地铁新技术界定、聚类分析和Logistic回归模型三个方面进行介绍，作为实证研究的理论基础。此外，文献综述部分将对以往研究中关于影响居民出行方式选择的普遍因素，公共交通新技术及模型应用进行梳理、分析和评价。

第三部分，地铁乘客行为模式挖掘。本部分主要利用Python和SPSS技术对已有数据进行数据处理和时空模式挖掘（聚类分析），从而在多个维度探究地铁新技术政策的提出对客流量的影响。

第四部分，构建模型。本部分通过建立Logistic回归模型、调节变量分析、控制变量分析和内生性因素分析来挖掘居民出行选择与地铁新技术性质之间的相对重要性及非线性关系。

第五部分，其他地铁新技术的统计描述。本部分将会再介绍地铁屏蔽门和地铁空调通风系统两种地铁新技术，利用现有的数据资源进行基本统计分析，进一步验证结论的普适性。

第六部分，相关政策建议。本部分根据上述地铁新技术对客流量的影响结果和居民出行选择与地铁新技术性质之间关系的研究结果，对后续地铁行业的科技发展方向和绿色出行实践提供合理化建议。

第七部分，研究结论。本部分基于上述的探讨与分析，得出了本文的总体研究结论。基于以上研究内容与研究步骤，本文的技术路线如下：

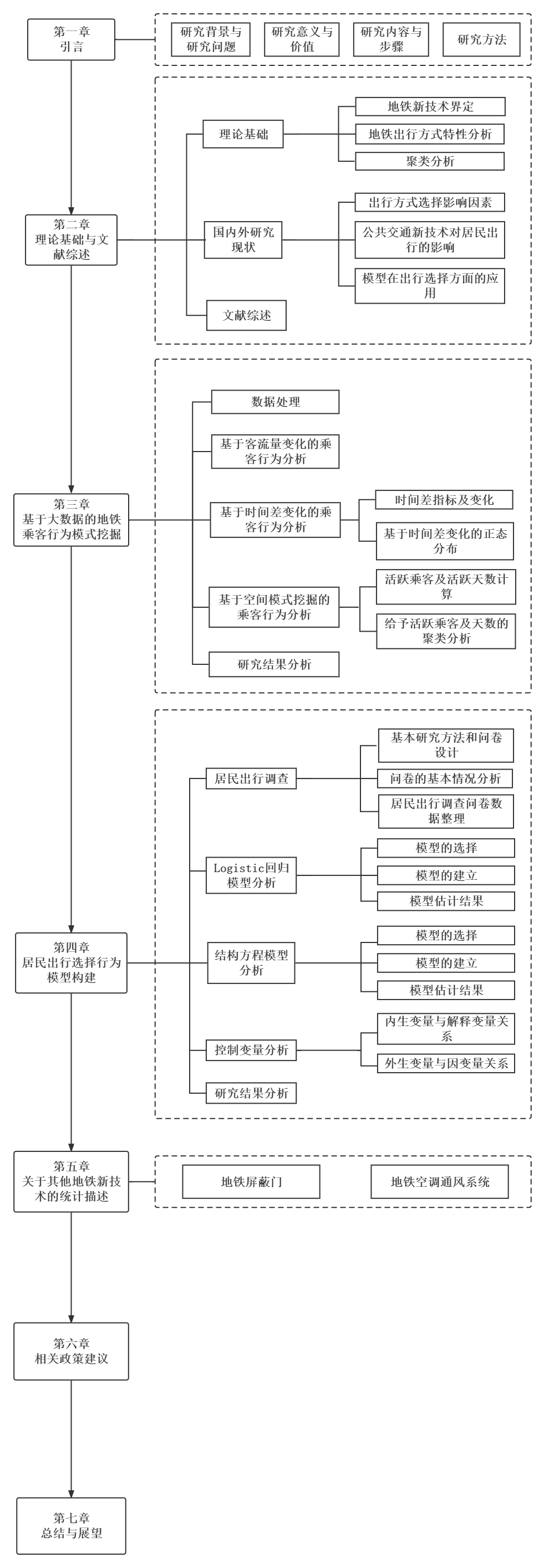


图1-1 研究技术路线图

* 1. 研究意义与价值
     1. 理论意义

（1）对居民出行而言。探究地铁新技术对于居民出行方式选择的影响是对影响居民出行选择因素中的态度、习惯机理、社会规范、技术属性及出行属性内容的重要补充。

（2）对绿色出行而言。倡导鼓励市民选择绿色出行的公共政策是政府部门必须落实的一个内容，本文的研究也能够完善环境领域相关的公共政策内容。

（3）对地铁行业而言。本文通过对地铁新技术与居民选择公共交通绿色出行行为关系的深入研究，来对传统居民出行行为加以扩展，不但弥补了居民出行行为在地铁行业的科研缺口[5]，而且还将未来出行行为领域与绿色环保领域之间构建起了密切联系，为绿色出行的实施提供了全新的研究视野和理论支撑。

（4）对科技变革而言。本文通过采用实证研究的方法来讨论地铁新技术的使用是否会对居民出行方式的选择行为产生影响，明确出居民对于互联网变革时代地铁新技术使用的接受和适应程度，这是对科技变革可行性分析的有力证据，并为接下来的轨道交通行业科技变革的方向以及“互联网+”技术的引入提供理论和方法的指导。

* + 1. 实践价值

（1）对社会而言。本文能够进一步证实互联网科技对地铁行业产生的积极影响，从而推动科技变革在居民出行方面的发展，增加居民出行的便利性和多样性。

（2）对政府而言。本文回应了环保部在《关于加快推动生活方式绿色化的意见》中指出的，不断推进绿色出行的发展，减少环境污染，构建低碳社会，促进可持续发展战略[6]。同时，政府可以在发布新的地铁新技术（与居民出行相关）使用政策之前，参考本文对该地铁新技术与居民出行选择关系进行研究和预测，从而保证政策的可行性和可靠性。

（3）对行业而言。本文能够积极推进并指导“互联网+”在轨道交通行业的应用方向，对地铁新技术的发展起到了积极影响。该研究结论也明确了科技和绿色生活之间存在的积极关系，对加强绿色出行乃至生活方式提供路径[7]，具有重要的战略意义和指导意义。

1. 理论基础与文献综述
   1. 理论基础
      1. 地铁新技术界定

地铁新技术主要包括地铁轨道线路结构新技术、城市地铁车站新技术、城市地铁载运工具新技术、城市地铁供电新技术、城市地铁通信信号及列车控制新技术、城市地铁系统设备新技术、城市地铁智能化系统新技术、城市地铁规划和设计新技术以及城市地铁运营组织管理新技术[8]。这些地铁新技术的设计路线可以总结为通过合理利用城市地下空间来实现城市轨道交通上下一体化，探索出城市轨道交通的可持续发展模式，以及推广更加绿色节能、高效便捷和先进创新的城市轨道交通理念。

由于本文的研究内容为探究地铁新技术与居民出行选择之间的关系，因此本文所提到的地铁新技术将锁定在与居民出行相关的城市地铁系统设备新技术、城市地铁智能化系统新技术、城市地铁规划和设计新技术以及城市地铁运营组织管理新技术这四个领域之中。然后再根据这四个领域内的地铁新技术与本文研究主题的相关性，进一步对地铁新技术进行选取和界定。

关于城市地铁系统设备新技术。传统的地铁系统设备包括AFC自动过闸检票机、地铁屏蔽门、地铁车站自动门、地铁空调通风系统等[8]。从运营角度来看，地铁的系统设备主要包括地铁运行系统、地铁服务信息系统和检修保障系统。其中，与本文研究内容相关的为客运服务系统。地铁服务系统设备新技术有AFC自动过闸检票机、乘客信息技术、无触点IC单程车票技术[8]。AFC票务清分系统是以AFC中央计算机系统为基础，能够作为城市地铁网中的发卡、清分数据、AFC运营管理的中心[8]。这个系统能够将各个线路地铁的重复票务以及系统参数设定进行共同管理分类，并且能对所有的交易数据进行整合分析并生成相应报表，最后对相应的收益进行清分，AFC票务中心系统还能够将地铁各线路之间以及和银联、各种移动支付运营商、城市APP等系统之间进行收益清算、数据处理整合以及统计分析并进行管理[8]。乘客信息系统是基于电力载波通讯技术的地铁应用。该系统的关键新技术为无线局域网应用技术及电力载波通信应用技术，可以让车站在任何状况下为乘客播放列车运行的即时咨询，为乘客咨询的多样性，互联网和标准化的发展打下基础[8]。无触点式IC单程车票的主要技术特性包括每单交易响应时间少于0.2s，平均故障低于0.1%，可进行500次交易，使用寿命至少半年以及具有一次性可写的储存空间[8]。该技术标准的提出克服了传统单程车票不环保、成本过高等的问题，并同时也为城市地铁电子收费管理系统的车票媒介载体设计提出了依据。

关于城市地铁智能化系统新技术。城市地铁智能控制系统主要分为综合监控智能系统、乘客信息智能系统、综合安全智能系统、通讯智能系统和信息智能系统。其中，与本文研究内容密切相关的即为乘客信息自动化控制系统[9]。该信息系统向乘客提供信息咨询服务，内容主要包含ATS循环报文信息、应急资讯、多媒体视讯信息和其他资讯等,还可以利用广告信息的发布来获取收益，并以此反过来推动信息系统建设[9]。车辆信息系统是以电脑为基础,并以车站和车载这一类显示终端装置为主体，利用多媒体互联网信息技术来向乘客们传递动态资讯。

关于城市地铁规划和设计新技术。该方法属于理论的新方法，其基础为分析影响因子、范围分析、地铁系统的区域性、研究方法和设计方法，主要内容包括城市和地铁系统特性两部分。其中，对城市规划产生的主要影响因素有自然环境条件、城市规模、城市人口、城市的经济基础、城市交通状况、土地利用程度及城市发展趋势。地铁系统特性包括系统的形式、运行方式和技术水平[10]。城市地铁规划的核心思想包括基础概念规划、地铁在城市布局中的作用、地铁的可持续性发展、交通疏堵的兼顾发展、整体计划的流动性和地铁各线路功能服务一体化，比如给车站设计屏蔽门或者在车站进出口设置更加便捷的二维码刷码过闸，打造以人为本的地铁及车站设计[10]。规划与设计的系统分析方法必须要将内外部条件进行结合，比如将目前利益与长远利益相结合，局部效应与整体效应相结合，定量分析与定性分析相结合。利用上述过程得到关于城市地铁规划和设计的新型理论技术作为未来地铁行业发展的指导依据。

关于地铁运营组织管理新技术。地铁运营组织管理新技术主要包括城市地铁运营组织管理传统技术、城市地铁一体化运营、城市地铁网络化运营、地铁安全运营技术以及城市地铁列车运行图编制系统[10]。传统技术包括客流分析、地铁车辆选型和编组、地铁线路计划、地铁停站计划、票务管理等。由于客流的不断变化以及科学技术的发展，政府需要设计出足够高效的运输计划，满足乘客的要求，构建以城市地铁系统为基础，其他系统为辅的多元化、多功能的城市地铁网。城市地铁交通的统一运营是指统筹好各个管理部门工作，使城市轨道交通各子系统间和外界因素的高度配合，各种公共资源都可以听从政府统一调度和管理，从而能进行整体迭代优化[11]。一体化运营将“以人为本”作为基点，使乘客能够便捷出行，减小城市交通压力。一体化运营的基本模式，包括共线经营模式和枢纽换乘模式。共线运营的硬件要求包括轨道技术标准一体化、站点设施一体化车以及车载运行工具一体化[11]。而共线营销的主要软件特点,则包括管理与协作机制统一、票务管理统一、物流组织统一,以及网络一体化。互联网运营以技术创新视为发展核心,主要包含系统创新、机制创新、体系革命和业务创新。安全运营技术主要包括总体研究、车门控制系统研究、地铁辅助逆变器启动失效分析，以及其新型传感器研发、电力调度操作票系统和设施安全及应急事故处理五部分内容[11]。其中与居民出行相关的为总体研究、车门系统研究以及设施安全及应急事故处理。总体研究要思考安全性和可靠性框架及模型，并进行指标体系分析评估。车门技术改善与优化工程主要涉及塞拉车门改装措施、内藏车门改装措施和外挂车门改装措施。

本文主要以城市地铁系统设备新技术中的二维码过闸机技术为主，地铁屏蔽门以及地铁空调通风系统为辅，来研究地铁新技术与居民出行选择的关系。由于地铁新技术的种类繁多且缺少一些新技术投入使用后的相关数据，所以本文只能作为该研究领域的一个开端。该领域的研究将会是一个循序渐进的漫长历程，而本文的研究过程可为后续对其他地铁新技术与居民出行选择关系的研究作为一个参考。

* + 1. 地铁出行方式特性分析

地铁通常是指利用轨道在城市地下运载乘客及货物的交通运输方式。第一条地铁于1863年在伦敦建成。从那之后各国的大城市都开始提倡城市轨道交通运输体系。地铁能够从根本上缓解都市交通拥堵、土地面积拥堵以及碳排放等问题，这些性能使地铁在较为发达的城市公共交通体系中逐渐变成主流[12]。地铁使现代科技变革以及全球经济发展共同作用的结果，更是满足了城市居民交通日益扩大的需求。地铁具备环保低能、节约地上空间，准时高效，舒适安全，可靠且运输量大等优点，可以满足乘客长距离出行以及大城市上班族需求的隔离性地下交通运输方式。当然，地铁的建造也存在一些约束性，例如城市范围较小，地下空间较小，技术不够先进且经济不够发达省市是不适用地铁且不容易构成轨道交通网络的。但是它身为一种容量较大，绿色环保的公共交通出行工具，仍然是公共交通未来的发展主要趋势。

地铁行业中存在了五个潜在变量特性因素能够影响居民出行决策。这些潜变量是根据之前的研究所构想、分析假定等客观存在的变量，本文也将利用这五个特性因素建立与居民出行相关的模型。

**（1）有形性**

有形性是指设备条件以及工作人员的体态特征。乘客会根据这些直观因素来评估服务质量。地铁行业也属于服务行业，因此需要依靠有形性来树立公司形象，为乘客树立对地铁行业的信心。有形性将会和质量维度联合来构建行业的服务质量。运载行业实质上不是一个具体化的产品而是整体运营的行为过程，因此乘客可以根据有形的角度来评价服务的质量[12]。

**（2）可靠性**

可靠性是地铁公司对乘客在质量舒适性、提供服务、解决问题、运送货物、准时性和确定价格等方面的承诺。乘客更倾向于选择能够遵守承诺的公司，尤其是那些以服务质量为核心的公司[13]。

**（3）响应性**

响应性通常是代表向顾客提供帮助来提高顾客对服务的满意意愿。该因素主要强调要高效快速的解决客户请求或者投诉等问题。响应性的主要以客户等待解决问题的时间来体现。相应性也包含能够满足客户要求的能力和灵活程度[13]。

**（4）保证性**

保证性是指企业的能力和工作人员学识及礼仪等方面的评估。在客户面对存在高风险的项目且本人无法进行预估时，该维度就显得尤为重要，比如在保险证券以及飞机医疗等行业。它能够提升乘客对地铁行业的信任度和安全感。在地铁中最好的体现为乘务人员以及地勤人员的业务能力和服务态度[14]。

**（5）移情性**

移情性通常是指给乘客提供个性化的服务。比如地铁行业中如果存在行动不便的患者，需要对这些患者提供特别服务。这将会面对更加广泛的乘客受众。

根据上述地铁新技术的核心理念，我们可以得出地铁新技术改善了设备条件、能够影响地铁的舒适程度和准时程度、更加快速的获取乘客问题数据并反馈以及增加乘客对地铁的安全感，因此可将有形性、可靠性、响应性、保证性和这四个维度作为地铁新技术的性质与居民是否选择乘坐地铁出行建立模型进行分析，来挖掘这二者之间的联系[14]。调查问卷（见附录）主要将这五个维度进行了具体描述以便参与人能够进行清晰评估。这四个维度共包括12项具体内容，如下表。

表2-1 地铁新技术属性及具体内容

|  |  |
| --- | --- |
| 属性 | 具体内容 |
| 有形性 | 二维码过闸机进出站满意程度  等候区屏蔽门满意程度  快速进站满意程度 |
| 可靠性 | 空调自动调温系统满意程度  减震满意程度  实时语音提示满意程度 |
| 响应性 | 多媒体视讯满意程度  地铁站位置的满意程度  换乘步行时间的满意程度 |
| 保证性 | 车门系统的满意程度  轨道行驶稳定性满意程度  时间准点的满意程度 |

* + 1. 聚类分析

聚类分析方法是利用确定一个中心将所研究对的对象数量加以划分的统计方式，一般用“间距”或者“相似系数”来衡量对象内部的类似或者相异性[15]。将这种类似或相异的信息看成各元素相互之间的“间距”长短的一个衡量，将间距较短的元素再细分为类，而不同的元素则间距就会很大[15]。聚类分析方法一般按照分类元素的数量不同可以分成Q型聚类分析和R型聚类分析, 对变量进行聚类分析方法的R型聚类分析，而对样本数量进行聚类分析法的为Q型聚类分析[15]。论文主题是基于大样本数量的Q型聚类分析。而将群集分类的方式, 主要包括了系统聚类法和动态聚类法。而因为本文中所研究的样本数量很多, 所以将会考虑选择在动态聚类法中的k-均值聚类法[15]。

聚类分析按照聚类算法分类包括层次聚类和划分聚类。层次聚类的基本过程是首先定义每个观察值，然后计算每个类别的聚类，接着将间距最短的两类合并为类别，通过不断合并来逐步减小类别的数量，直至将每个观察值的种类迭代为下一个类别为止[16]。在层次聚类算法中，主要区别是度量类与类之间的距离方法。

表2-2 层次聚类方法[16]

|  |  |  |
| --- | --- | --- |
| 聚类方法 | 两类之间的距离定义 | 备注 |
| 单联动 | 一个类中的点与另一个类中的点的最小距离 | 适用于细长的类别 |
| 全联动 | 一个类中的点与另一个类中的点的最大距离 | 适用于相似半径的紧凑类别，能够感知出异常值 |
| 平均联动 | 一个类中的点与另一个类中的点的平均距离 | 适用于聚合方差小的类别 |
| 质心 | 两类中变量均值向量的距离 | 无法感知异常值，但能够削弱表现 |
| Ward法 | 两个类之间所有变量的方差分析的平方和 | 适合于仅聚合少量值、类别数接近数据点数目的情况 |

当需要嵌套聚类方法和有意义的层次结构时，层次聚类分析法比较合适，不过层次聚类分析法无法使用在成百上千的大样本观测值中。由于本文最后要处理的数据量高达8千万条，因此我们本文选择划分聚类的方法。划分聚类主要包括k-means均值划分聚类、Mini Batch k-means、k-means++聚类、bi-kmeans算法、kernel k-means和PAM[16]。

表2-3 划分聚类方法[16]

|  |  |  |
| --- | --- | --- |
| 聚类方法 | 优点 | 缺点 |
| k-means均值划分聚类 | 能够处理比层次聚类更大的数据集，对团状数据点集区分度好 | 可能会被异常值影响 |
| Mini Batch k-means | 处理较大数据集时，时间较短，算法收敛时间更短 | 准确度比标准算法更低 |
| k-means++聚类 | 有更高的概率选择当前离n个聚类中心较远的点作为新中心 | 改进过程较为麻烦 |
| bi-kmeans算法 | 由于SSE最小化原理，SSE可以成为衡量聚类分析法效率的重要指标，改进了k-means算法会陷入局部最优的缺陷 | 对簇进行分类时需要考虑如何能够最大程度的减小SSE的值 |
| kernel k-means | 非线性映射能够提高数据点线性可分的可能性，甚至在聚类算法失效的情形下，通过引入核函数也能够获得更加精确的聚类分析效果 | 非凸类型的数据簇在现实生活中并不普遍 |
| PAM | 对比之前的均值聚类法更加稳健，可以使用任意的距离来计算 | 重复步骤较多 |

k-means法（也称快速聚类）计算的工作量较小，但计算效率却比系统聚类法高。k-means法并不是必须要提供所有可能结果，而仅仅要求学者们先把将要界定的类型个数定义起来，接着再界定各个子聚类中心[16]。接着,再测定各样本与聚类区的间距,并根据距离的长短加以分组。“k”就是代表类别个数，均值则代表聚类中心。具体步骤如下：

（1）确定类别个数k。

（2）定义k个初始聚类中心。要求选择k个样本作为k个类别的初始聚类中心。

（3）按照已确定的k个初始聚类中心，计算每个样本到k个聚类中心的欧式距离，并按照最近的原则将所有的样本划分到预先设定的k个类别内[17]。欧式距离公式如下式2.1[17]：

dij(2)= （2-1）

（4）通过k个类别，统计出各类别中每个变量的均值，并以均值点作为新的k个分类中心。通过新的聚类中心，可以重新估计每个样本到新中心的距离并重新分类[17]。

（5）重复以上步骤，直至达到满足终止聚类条件。终止聚类必要条件是：a.迭代次数必须达到事先规定次数。b.新的聚类中心与上一个迭代的中心点的最大偏移量小于指定量[17]。

* 1. 国内外研究现状
     1. 出行方式选择影响因素

从20世纪70年代起，国内研究者逐步地把环境心理学、社会学、市场管理学、市场营销学等研究领域运用到对城市居民出行方式选择的研究中。在这个阶段主要经过了三个时期, 20世纪70年代的研究阶段重点在于探讨各种个人的社会心理性变化及其对城市出行方式选择的作用，此方向的研究成果则主要采取了单纯的量化或定性分析方法[18]。而20世纪90年代则把家庭因素当作行为的理论基础，认为城市居民的出行方式决定是一个社会发展过程。个人都在不断的比较自己与周围群体行为上的共同点与不同，从而利用行为选择或模仿来保持或改善自己固有的出行模式，也因此社会规范、社会认知等群体性因子也将制约着公民出行模式的选择，本领域的研究重点主要以定性分析技术为首[18]。于2000年根据人类理性行为特征，分别探讨了价格、技能、天气、政治等社会情境因子，以及人群、地域、社会经济条件等结构因子对公民出行模式产生的影响的相互作用问题，其分析方式主要采取了定量分析为先、相对定性为辅的多元分析法[18]。

国内研究者对关于旅行方式的心理影响因素的研究重点主要围绕着内在心理性因素、外在情境因素和结构心理因素三方面：

（1） 心理因素

环境责任心是指个人对做出特定环保活动时的负责与道德的心理意识。而关于个人环境责任心及其对个人环境友好行为的影响，目前国内已有研究成果均获得了较一致的正面与相关结论。1978年，Dunlap等研究者表明，环境责任心与能合理预测环境行为，两者之间具有正向相关的关系[18]。Hines等人给出了社会责任的环境行为模式, 研究结果表明，通常拥有较高环境责任心的个人更易于参与环境友好活动，同时环境责任心也通过行为意愿间接影响了环境友好活动，当环境责任心扩展至关于绿色消费行为的研究时，结果也一致[19]。王建明和王俊豪在调研时认为，环保责任感是影响消费者低碳生活方式的关键因子[19]。岳婷等通过扎根研究和对城市居民的节约行动研究中表明，环保责任心对城市居民节约行动具有重大影响关系[19]。谢守红等通过实验研究表明，环保责任心是影响城市居民低碳消费的主要原因[19]。

在对出行方式影响的研究中，祝伟等在对环境影响因素的调查中加入了环境认知变量，从而证实了环保认知能够影响市民的出行方式选择[19]。王世通与陈君彦合作运用结构方程模拟研究了天津市城乡居民的低碳出行行为，证明了低碳出行的行为知识也会间接影响低碳出行活动[19]。潘静玉的调查研究也表明，大量的环保知识促进了市民对绿色出行行为意愿的产生，从而影响行为的产生[19]。Delis和Iosifidi调查也指出由于个人和家庭的环保意识不断增强，导致了私人车辆使用的频次下降，同时使用公共交通的美国市民数量也翻了一番[19]。Anable等把环境意识界定为个人对环境改变的实际(需求意识)及其自己生活方法所导致的社会环境后果(后果意识)的科学而全面的理解，个人尽管具备必要的环境意识，但如果不相信自己的环保行动将引起社会生活环境的显着性变化，就无法实施环境行动[19]。Schwartz和Howard提倡的个人规范决策模式，认为对行动所需要的自由意志、对行动结果的意识，是影响个人规范的先决条件。Antimova等人认为，只有在普通市民意识到小轿车的使用将会损害环保，并觉得自己应当负有责任之时，个人规范意识才能真正被激发，进而人们才会选择更加环保的出行方式[19]。Joireman等在调查中认为， 一些更偏爱公交的市民们的确具有更强的需求意识和后果认知意识，同时他们也更注重于未来的环保变化，并认为小轿车的使用对环保的损害是长久性的[19]。Yang和Long利用对江苏省城镇居民的调查研究表明，个人对环境危机意识显著影响了居民个人对都市公共自行车服务项目的投入意向[19]。

环保责任感是指个人对于做出一个环保活动的责任与道德的心理认识。关于环境责任心对环境友好行为的影响，目前国内已有研究成果都获得了较一致的正面与相关结论。最早的1978年由Dunlap等学者的研究表明，环境责任心与能合理预测环境污染行为，两者之间具有正面且相应的关系[19]。最具代表性的是由Hines等人提出负责任的环境行为模式, 研究成果表明通常拥有环保责任心的个人更易于参与环境友好活动，同时环保责任心透过行为意愿间接影响环境友好活动，将环保责任心扩展至关于绿色消费的调查时，亦得到了相当一致的结论[19]。Wilhelmsson通过把家庭的社区文化特点和环境知识变量带入文化特征消费过程中考虑消费者对绿色住房的价格支出状况，表明文化收入特征和环境知识对家庭在绿色住房的消费行为上具有很重要的影响[21]。王建明和王俊豪在分析中都认为，环保责任感是制约中国消费者低碳消费模式的关键因子[21]。岳婷等人用环境扎根理论在对城市居民的节约行动分析中表明，环境责任感对城镇居民的节约行动具有重大影响关系[21]。谢守红教授等人经过实证分析后认为，环保责任感才是影响中国城乡居民低碳消费习惯的最主要原因[21]。

态度是指个人对实施某种特定行动时的主动或消极程度的评价，在环境行为的研究中，研究者们主要偏向于把社会环境心态当作对社会环境行动的主要影响因素[21]。在社会生活方法的研究中，研究者们的研究成果则主要偏向于把社会环境心态分成了感知性心态和情感性心态二种层次[21]。对认知性心态的评估又从二个方面出发，首先是在充分考虑生产成本、收益原因后权衡出行方式是不是必要(生产成本心态)，第二次是在充分考虑环保因素后权衡出行方式是不是意义(环境态度) [21]。Black等人认为相较高约束的环境行为，认知性态度对低约束的环境行为具有更强的理解能力，而低约束的环境行为则多指一些较低成本的简单活动，如在旅行方面的选择[21]。Vining和Ebreo研究结果表明，在环境态度中的思维和情感因素可以增强人们对环境行为的预期[21]。动机通常决定于个人的心态、想法与行动。Geng等研究表明居民的外出行为必然会伴随着各种动因或偏好，而只有主要动因才是影响居民日常外出行为的最关键因素[21]。在此基础上，Geng等人通过动因与行为的双维视角，建立了三种色彩(四种类型)的出游行为结构模式，包括了绿色出游行为，红色出游行为，无奈式灰色出游行为以及条件式灰色出游行为[21]。而经过实验调研后表明，环保动机是绿色出游行为的必需而不是充分条件，从而维护了绿色出游行为的稳定性，其安全性、经济效益、舒适度或便利性都是造成出现动机与行为不统一现象的关键因素，从而调节、影响并改善了市民的出游行为[21]。Geng等的信息干预实践研究表明，拥有不同目标框架(动机)的群体，其对信息干预的反应存在显著差异，其中以环保动机作为主要动机的被试群体，在信息干预后更多地使用了绿色交通工具(步行、自行车或公共交通)来取代了私家车[21]。而将健康行动视为最主要动因的被试人群中，其选择绿色交通工具的情况则相应较低[21]。在绿色消费方面，期望是消费者的一个心理期望，是消费者产生愿望的一个重要心理基础[21]。对于满足期望值，消费者就会产生一系列的购物动机。所以，消费者期望值也会影响消费者的支付意愿[21]。黄海龙结合了绿色居室，研究结果表明消费者的希望能够从以下多个方面加以理解，即是否想有卫生、舒心的生活起居环境，是否希望为改变环境污染而做出自身的贡献，是否希望自己的居室相对环境污染较小，是否希望自己的居室使用、维护费用较少[21]。

习惯是由于出行行为本身伴随着大量的重复性。习惯也是一个非规范动机，会介入到个人行为决策流程中，在一定程度上影响规范的激活过程[21]。如果习惯强度够大，个体在行为决策时会大大减少对相关情境信息的搜索和使用过程，减少甚至跳过规范的激活和作用过程[21]。Klockner与Matthies认为习惯能够调整个人规范和日常生活模式间的关联，并且调整的范围和行为的好坏有直接关联[21]。De Pelsmacker和Janssens在对交通安全行为的分析中，都认为习惯与感知行为控制系统共同影响着行为意向与动作，不过二者解释的内涵却有所不同，因为习惯造成的是偏向机械的行为，而认知行为控制则造成的是偏向理性的行为[21]。

（2） 情境因素

Tajfel把认知性(identity)界定为“个人自身定义的部分，源于个体成为一种社会群体组织成员的意识，还有其与该组织成员身分或地位有关的社会价值和人文情感含义[21]。” 所以，认识或多或少源于个人对自己所处的内群体与外群体相互之间的利益对比。社会角色学说认为认同固然是个人对自我的认知，但这个认同更多的是经由与别人的作用所产生的，而这个作用是在个人不断地比较自己和别人间的共同点与不同。Breakwell的认同论具有以下三个假设：个人希望被社会认可、个人具有认同的连续性、个人既渴望多样性，也追求相似性[21]。随着全球文明的多元发展，面子作为一个中华民族文明积淀的产物，也越来越成为本土以及国内外专家学者研究的一个焦点。国外研究者比较偏向于把面子定义为某种文化或社区特定的现象，而对面子的理解也主要是基于身分、地位、名望、尊严等的某种社会认同感，提出面子定义应该与认同有着比较密切的联系[21]。不过，Tajfel与Goffeman则认为虽然二者并不是混为一谈，但因为社区认同感只是个人自身定义的一个，作为个人所感受到的“他们眼中的自己”，面子则是一个必须经由他们确定的、内部的社会自我价值，是由个人“期望道德自己”和别人评价以及回应相互之间的冲突所产生的心灵感受[21]。刘继富教授指出，面子是指个人藉其行动及社会化资源体现其自我人生价值，或谋求别人的肯定而获得意外的认可时，并凸显于个人内在的自我价值感及相应感受[21]。面子实质上是指个人对自己在别人眼中的社会价值和地位的关注，自身价值是面子的核心，而社会性资源则是个人体面的象征。郭晓琳和林德荣按照“心态—行动” 思维，从人们的面子起源与认识、面子消费行为的动因与策略、面子对消费行为的影响等三方面，汇总了调查状况[21]。汪涛与张琴把消费群体的面子认知机理分成了商品属性测量和参考市场客观评价，调查结果指出商品地位重要性越大，消费群体所认知的商品面子浓度就越大，而消费群体的面子要求也在当中起着调节影响，与较低面子要求的消费群体一样，对面子要求更高的消费群体所认知的商品面子含量也就更大[21]。

Ajzenh和Driver指出，社区标准即个人在行动中所感受到的来源于别人、团体或社区的压力，并受到社会规则信念与顺从规则的动力二部分决定[21]。中国学者郑晓明教授认为，现代社会规定是指一个社区和所有社区组织以及人员之间相应的行为、法规、惯例、道德和价值准则[21]。但大多数学者的研究都指出，社区规定会直接影响个人的资源节约情况，节能行为和低碳消费行为。Cialdini等把社会准则界定为弱于权威性的，用以规范或约束社会成员活动的规范与要求，并把社会准则分成两种：描述性研究法准则与指令性标准，而描述性规则对行为的作用更类似于从众行为的产生，而命令式规范则由社会评价体系对行为加以评价[21]。大多数的学者研究成果都表明，社区标准会直接或间接地影响个人的资源节省行动，节约行动，以及低碳消费。例如，Nolan等人采用电话问卷式调查的方法，对美国加州地区的八百一十名市民所开展的问卷调查，研究结果证明了社会规范对节能行为有着明显地作用[21]。王建明的扎根研究认为，社会规范对个人的节约行动的意义与行动，具有调节作用。芈凌云教授经过实验计算后得出，社会规范对城乡居民的低碳行为需求和对城乡居民低炭化能源消费四大层面上的行动(购置环保电器行动、购置绿色能源行动、住房节约投入行动和低碳化能源应用行动)均具有明显的正向调节作用[21]。张根林基于对重庆地区顾客对绿化心态的调查，发现了影响顾客选购绿化心态的主要原因是绿化的经济价值、消费者选择绿化的便利程序、包括顾客自己的卫生意识水平、环保意识和社区标准[21]。邱重植教授等人通过对中国浙江省城市和乡镇地区消费群体对蔬菜安全问题的态度、认识及其购物行为数据分析，得出了社会规范等情境因素影响下对健康购物行为产生的间接影响[21]。

耿纪超等人认为：社会规则与认知虽有相似之处，但也明显不同，因为社会规则即个人所感受到的源自于一种外在参照物的社会压力，而认知则体现为个人在这些社会压力下，在多大程度上愿意接受并服从于这个常态，从而把这个一致性体现了出来[21]。Kollmuss和Agyeman都认为：个人如何采取新的行动往往会引起对别人的道德影响，而一旦周围人都实施了某特定行动，那么该行动就通常会被看作是合理的，也因此大多数个人都在按照社会大众的道德价值观来不断地调整自己的道德准则[21]。假设周围人因为要求更高的生活品质而继续使用小轿车，因为个人很难通过自己的道德规范去批评身边人的做法都是不正常的[21]。相反，个人单方面的环保行动很可能会被社会批评为 “假正经” 或者 “另类”，而此时社会规范就发生了巨大的影响[21]。在出游方式选择的调查中，潘静玉、陈凯研究表明社会规范是影响居民出游方式的最主要原因，而当社会规范与个人规范并存时，社会规范的影响也可能会减弱或者不明显，因此社会准则内涵的部分也可以被个人规范所包含[21]。Donald则认为社会规范与个人规范关系的焦点有所不同，只是当二者发生冲突时，一般居民都会同时权衡二种原因，而社会规范则更多地作为调节原因而影响了个人规范，从而可以保证低碳出行方案选择的稳定性[21]。Klockner和Friedrichsmeier通过增加个人态度、愿望和习惯因素，来理解学生的出行行为，研究结果表明个人标准、社会规范和态度共同决定意愿，而愿望和习惯则共同决定了出行行为[21]。杨冉冉的研究还指出社会规范作为情境变量，对绿色出行活动产生了反向的抑制作用，即对低社会标准的个人来说，绿色出行活动愿望和行动间的正面影响相对较强[21]。所以，通过培养较低社会地位规范个体的绿色出行活动意愿，能更有效推动该个体的绿色出行活动。

制度技术情境主要强调除了心理影响以外的客观情境影响，主要涉及科技属性(交通工具属性、交通运输基础建设、信息技术、区位、都市结构)、旅行属性(出游时机、出游远近、出游目的)及政府有关规定等[21]。从而增强公交的技术属性，会促进市民的低碳出行方式。Van Vugt等调查表明，如果乘坐公交旅行的耗时会能缩短到一种理想的程度，市民对公交的选择意愿会明显提高[21]。Yang等人认为如果出行距离在五公里以内，则市民对小轿车的依赖性很高，唯有在出行距离在十至二十公里之间时，对公交的替代性才会明显提高[21]。以柴彦威为代表的研究者通过从人类地理的角度，指出了居民的生活特征并不仅仅依赖于时间与距离上的二维特征，还包含了空间特征，如区域、空间结构等。住户的集体意识与偏好也会造成活动结果的空间结构不同，而这些空间结构特性也会约束到住户的活动行为，如机动化对住户活动空间形成的深远影响，而且活动空间也与能否采用小汽车出游等存在明显的相互关系[21]。堵车会干扰市民的日常生活，而张小宁通过对比研究认为，若能形成完备的网络系统，让司机获取正确的道路数据，那么交通拥堵现象将减少。柴彦威、甄峰等专家提出在大数据处理的新时代背景下，计算机技术也可以和人文地理相结合，利用大数据的获取、挖掘、压缩、流动，来促进人们行为在大空间中发生改变[21]。另外，根据klockner和Friedrichsmeier研究表明：汽车出行的天气条件、气候类型和旅行的目的都显著影响着城市居民的出行方法，在下雨天和相对冷的天气，以及在一些以消遣娱乐为目的的家庭或活动，城市居民对小轿车的使用频次都会明显增多[21]。

（3） 结构因素

黄树森教授等人经过研究，认为年龄、职业、小轿车保有量等是影响人们旅行方式的最主要原因。PriLlwitz和Barr,lopezmosquera等人以及Clark等人研究表明，妇女、中年轻人、较高收入以及有更少私家车的城市居民，更喜欢选择徒步或公共交通等环境友好型的旅行方式，而其中百分之五十八的家庭中没有小孩,家庭中小孩数量越多，家庭中对小轿车的依赖度就越大[21]。马军红和廖娜等关于西安市城乡居民低碳生活情况的调查结果表明，人均收入越高、教育较好的人群对小汽车的依赖性也较高，而且这部分人群的低碳意识和低碳行动的可塑性也更高[21]。Susilo等人经过对英国社区住户的调查后表明,如果持有小轿车并不会危害住户的走路方式，但却会危害对公交的选用，住户如果持有小轿车，则其选用自行车外出的几率也将会大大降低[21]。Smart在对全美人口地理编码数据集的大数据分析中发现，新移入加州的美国居民较本土居民更偏向于采取低碳的生活方法，比如徒步、单车、公交等，而假设其同胞都是移民者，则这样对环境生活方法的选取情况就会比较明显[21]。Duarte等人运用了法国在2008年的社会经济普查数据建立了一种静态的CGE模式，随后又通过对微观的消费者行为问卷式调查数据研究，指出对消费导向所带来的碳减排效果在低收入群体中更加突出，针对汽车交通部门的消费导向是最有效的，但对食物部门的消费导向虽然会造成预期的甲烷排放量降低，但也会造成温室气体和二氧化硫排放量的上升[21]。Wolf和Moser分析了全球一百一十九个国家的社会人口学特点、地域特色等对我国公民气候变迁意识的相对影响，研究结果显示，在世界范围内，受高等教育程度是解释公民气候变迁意识的最重要原因[21]。但在国家的尺度上，影响因素也具有相当的多样性。因此，在国内最重要的因素是受教育水平、地域区位(城/乡)和家庭收入水平，而在美国最重要的因素则是婚姻状况和受高等教育水平。另外，新生活方式也是指个人或族群在消费、工作与娱乐上显示出来的特殊的行动模式以及目标倾向，它类似于传统习惯，却同时也涵盖了传统习惯以外的人口、地域与心理等特征。生活方式问题的提法来源于市场营销学理论的市场分类，它指出不同城市居民间的生活水平是具有差别的，但在同样的经济环境和社会条件下，许多居民都会产生了趋于相同的价值观或习惯，从而产生出了拥有不同生活方式的人群，且居民总是偏向于采取与他们的生存标准和喜好相符的行为方式[21]。例如，芈凌云，Nilssen研究表明人们在进行节能或低碳化行为时，一个相当重要的心理障碍就是人们对生活舒适度的要求[21]。假如城市居民相信节约或低碳化的能源消费将威胁到自身的生存品质，那么他们通常会对环保行为保持犹豫甚至质疑的心态，城市居民对日常生活舒适度要求的持续提高也将造成其能源消耗和碳排放量的持续和不断上升[21]。另外，Seligman等人利用这二项试验来调查城市居民的用电行为，而调查结果也说明除去个人舒适度的喜好之外，他们对身体健康、省钱、能量危机、以及科学信念的重视程度或倾向，也是解释其用电消费行为的关键原因[21]。在旅行方面的研究中，研究者们比较多地采用了全生命周期角度，主要关注于城市居民在休闲消费领域中的行动模式或倾向，认为许多城市居民选择和使用小轿车并不仅仅是为了满足交通运输的需要，更主要的是为了符合他们对闲适和满足他们日常生活水平的目标要求。例如，Lanzendorf就认为日常生活方式，是反映人们居住生活各方面都的行为模式，决定着居民们对自己所在的特定人群的潜意识的理解，但这种认同也可能迎合了普通市民在出行方式选择中的心理需求和社交要求，比如彰显身份地位、崇尚时髦等[21]。

* + 1. 公共交通新技术对居民出行的影响

由于目前关于公共交通新技术的相关介绍很少，所以在这里主要介绍两种资料较为完全的公共交通新技术。

首先是关于智能交通系统(ITS)的研究现状介绍。智慧交通系统是指交运技术在传统智能基础上发展到的一项更高水平，即智能阶段[30]。其实质是指通过技术对传统的交运体系加以改革，而产生的一套信息化、智能、社会化的全新现代交通运输体系，使交运基础设施充分地发挥出最大的效率，进一步提升服务水平[30]。ITS的最初概念是由美国政府于20世纪60年代发明的，开始时候叫做智能汽车交通控制系统。美国的智慧公交系统计划重点是基于移动公交系统的即时调度技术和即时数据的技术，以及利用新型的电子、通信信息技术改善公共交通质量与服务的应用方法[30]。具体内容涉及：车辆管理、出行者信息、电子收费和道路需求信息管理等多个方面[30]。其中，车辆管理所重点研发通信系统、地理信息系统、自动汽车定位、自动乘客计数、交通运营管理软件，以及交通信号优先[30]。而出行者信息系统则重点研发出行前、在途的公共信息服务管理系统，以及通过各种出行方式的接驳公共信息服务管理系统[30]。日本的ITS起步相对较晚，但受到政府关注，其发展与引进的速度也非常快[30]。日本的城市智慧化公交系统发展经过了3个发展阶段：20世纪70时代末已开启使用公交定位管理系统—公交车接近显示管理系统，80时代初期已开启使用公交运营管理系统，当中涉及乘客的自动计算、运营监控和服务管理，90时代初期已开启使用公共交通服务综合管理系统，当中涉及后勤服务管理信息系统和运营支持系统[30]。1985年，欧共体内十九个国家为主的政府部门和民间企业机构经过联合后，共同推动了ITS的发展规划，并改名为欧盟交通运输技术实施机构，共开发投资约五十亿元，并制定 “欧盟用作车辆安全的专业道路基础设施规划”，重点研究高速公路交通需求控制、交通运输与旅行信息、市镇与城市综合交通、辅助驾驶技术、物流与企业车辆管理、公众交通管理技术等，此规划于1994年实现[30]。韩国政府的ITS示范工程项目选择了在光州，该工程项目中选择了汽车道路感知信息管理系统、公交客运信息系统、汽车动态线性引导控制系统、汽车智能化管理系统、信息播报管理系统、电子收费管理系统、停车预警管理系统、汽车运行中的测重控制系统[30]。新加坡目前已具备了现代化的城市系统，该管理系统除城市控制系统中传统的功能如信息管理、道路交通监测、交通引导和道路交通资讯服务等之外，还包括了使用电子道路收费卡管理车流量[30]。20世纪90年代初期开始，中国国内外的部分省市在使用与推广ITS管理系统方面也获得了一定的成效，如目前电子商务收费管理系统的使用范围就有全国数十个省市[30]。我国目前已经在杭州、广州、北京等地安装了电子车牌,和汽车GPS定位设备，并实现了对汽车的实时跟踪、自动定位、班车与调度室之间的双向通信。广州己经制定了发展ITS体系的一五年计划，将建设成包括广州市范围内道路及交通信息资源共享体系的主要网络平台、城市交通及公共信息服务网络平台、城市间静态道路及交通信息网络系统等在内的服务主体框架[30]。目前，除了出租车的GPS系统以外，广州市交通委也正积极推进一批信息化交通工程。广州的 “公交GIS系统” 以及公车免费信息电子触摸屏的推广速度也较快[30]。ITS技术对中国城市公交系统中的运用还处于起步阶段，但随着进展速度极快，该项技术在全国大中城市公交体系中的运用将会越来越多。

首先是关于二维码支付在公共交通的使用以及二维码刷码过闸技术研究现状介绍。国外对二维码技术的探索也起步于20世纪80年代，但二维码在中国城市的经济发展和市民生活方面都有着广泛的影响[30]。通常会用到企业、运输人员对货物的配送处理方面。海外对微信二维码技术的研发晚于中国，在现实的环境中，手机扫码消费并不普遍，促使国外发展更多的是刷卡付款模式[30]。二维码技术的研发在我国起步于上世纪九十年代， 自2011年开始，二维码的应用也迎来了大爆发[30]。支付宝手机app上线二维码支付服务，而2013年微信5.0版又增加了二维码付款功能[30]。代表阿里的支付宝和腾讯微信两大移动支付公司将手机二维码支付技术带到了全新的高度。人们对二维码支付技术在交通运输的使用上提出了巨大的要求，紧接着我国交通运输部将颁布了《交通一卡通二维码支付技术要求》[30]。目前，中国国内各种地铁的产品都全部使用了二维码技术来进行交易的。二维码技术主要使用在购买车票、出闸乘车业务和日常的车票办理等业务。在中国各地，城市轨道交通领域应用最普及的二维码支付方式主要有网络车票、行业二维乘车码、微信/支付宝等二次开发的乘车码三种使用模式[30]。网络车票使用时，旅客须提前安装相对应城市地铁的网络售票app并进行个人验证，同时绑定支付账号，在上车前使用该app和第三方付款服务订购单程的网络电子车票，由互联网售验票平台在审核并支付后，向旅客的手机app上发出二维码电子乘车凭证，旅客持该凭证直接在城市轨道交通站点二维码刷码出闸，由互联网售验票平台在证明网络查票机上所获得的电子乘车凭证为合法有效后允许验票机开闸放行，对非法或无效电子乘车凭证的管理将仿照单程票规定统一办理[30]。行业二维乘车码服务(ODA)是指采用ODA后付费技术，采用的双脱机及信用付款模式，该模式与传统互联网电车票的最大不同之处是先上车，后付费业务简单，且无需受手机终端联网的影响，且通行效率高，稳定性较强，是目前中国各大中城市地区轨道交通业务中优选的手机二维码付款方法[30]。而二维码标准规范对二维码数据结构进行了定义，设计了一种混合码编码方式。它由行业数据域和客户数据域所构成。地下铁路方通过地铁AFC统一的自建3DES对称加密算法来保障安全，将用户数据域通过与手机二维码合作的发行方式分发，并通过SM2非对称国密方法来保障安全[30]。手机二维码架构设计更加灵活，提供通用乘车码和出进站分离乘车码二种发码方法，并允许通过多种支持方式按统一标准连接南京地铁的网络售票体系[30]。

* + 1. 模型在出行选择方面的应用

本文主要使用Logistic回归模型和结构方程模型来对居民是否选择乘坐地铁出行进行研究，所以在这里主要介绍非集计模型和结构方程模型在出行选择方面应用的研究现状。

首先是关于非集计模型在出行选择方面的应用。在20世纪70年代后期，由Manheim、Ben-Akiva和Lemman研究小组，初次将经济学里的效用论从应用范畴延伸进入了交通运输范畴，同时又以概率观点为基础，从非集计的理论层面上开始人们对交通运输方式选择的行为问题开始了研究，也因此将非集计模式带到了实际应用的强层面上，从而使非集计模式获得了进一步的发展完善，从而产生了在理论层面上研究成果更为丰硕和更复杂的改进型的非集计模式[36]。例如，Chu开发的PCL(Paired Combinatorial Logit)模型，Bunch开发的 MNP(Multinomial Probit)模型，Vovsha年开发的 CNL 模型(Cross-Nested Logit)等[36]。随着多个省市的居民出行研究以及一系列出行情况特殊研究的开展, 在调研统计的基础上，对个体交通方案选择情况的深入研究, 取得了大批的成果。Grayson，Wilson，Forinash和Koppelman，以及Bhat等人已先后使用非集计模型，系统分析了个体旅行的交通方法、旅行目的及其出游行为特点[36]。在多项Logit模型(Multinomial Logit, MNL)的基础上，Nested Logit模型于上世纪八十年代初期在信息交通领域内得到了迅速的发展，在多种决策步骤之间存在着阶层关联的选择中获得了广泛运用[36]。基于工作出行者对城市交通活动的复杂性，利用了多项Logit(MNL)的模拟调研研究了城镇居民工作出行方式的选择情况，并把核心课题放在了换乘因素对城镇居民工作出行方式选择的影响上的研究分析[36]。Tu T.Ton和DAvid A.Hensher共同提供了使用BP人工神经网络技术分层并结合分层的Logit模型，来处理城市交通方式规划问题[36]。Chandra通过对汽车个人偏好和属性中的无法测量和可衡量变量的分析，提出了基于多项Logit模型来刻画人们对汽车旅行方面的个人偏好，以及对个人在不相同服务水平下的交通工具选择情况做出了分析[36]。同年，Palma等人利用NL模型，调查了日内瓦市民在上班与外出时的交通方式选择情况，调查结果表明，如何选择小汽车也是一个影响出行者交通选择的重要要素。Cherchi等人通过RP/SP研究，构建了在意大利Cagliari市的居民出行巢式Logit(NL)模型[36]。Giulio Erberto Cantarella， Stefano de Luca阐述了运用MLFFN (multilayer feed forward network)方式分析人们出行需求中的重要环节，并清晰地介绍了这种方式还较好地应用在生活方式选择研究中[36]。Stephan Krygsman与Theo Arentze等用经过改进后的MNL模型，研究了出游方式选择和出游项目选择之间的相互联系，研究结果表明针对不相同的出游项目，出行方式与选择结果的差异比较明显[36]。总之，建立和应用对这部分出行特征灵敏的模型是十分必要的，这些特征体现着个人交通管理方法的选择。当前建立起来的非集计模型中还有许多不同的类型，例如Logit模型、Probit模型或者Logit模型的改进模型等等。最先被提及的大都是Logit模型，非集计模型中最具有一般性的就是Logit模型，由于其表现形式相对简单，也比较容易使用，不过Logit模型又有其本身的局限，也即它的随机效用项必须是可以单独确定的，所以许多研究者就一直在努力着对Logit模型加以创新[36]。但是，这并无法从根源上解决掉Logit模型的缺陷。其他的模式如probit模型尽管在一些程度上避免了这种缺陷，但是这种建模方法却需要借助于非常复杂的Monte-Carl仿真方法和多项式Clark求解逼近计算。想要克服上述难题中的一些缺点，还必须对交通方法的选取过程做出更深入的剖析，从而使得这类方法更具有实现性[36]。

对于居民交通方式选择行为的研究，中国在该领域的众多研究人员也都投入了大量的分析。对城市居民交通方式选择的主要影响要素的深入研究，能够为城市交通计划，特别是发展公共交通的长期计划提出意见，并可以更有效地引导城市居民交通的方式选择。隽志才对人类依据活动链的城市交通需求预测进行了主要定量分析，对人类社会经济活动和人们出行行为中的重要决策特征做出了主要定量分析，并深入分析了MNL模型和NL模式在城市交通运输需求预测领域中的实际运用[36]。王树盛等人认为Mixed Logit模式和Logit模型比较，具有更丰富的物理内涵，对于个人出行及交通模式选择情况的描述，也更加贴合社会现实[36]。邵昀泓提供了一个通过基于效能最大化概念的多项Logit(MNL)模型对个人公共交通活动做出决策研究的途径，并建立了个人对公共交通活动选择可能性与出行属性等因素间的类似线性关系[36]。郭寒英使用了多目标式决定方法、效益最大化理论和Logit模型方法，建立了旅行者的交通学习方式与选择模式等经济价值研究模型[36]。陈团生利用对北京居民生活的调研与数据分析，按照非集计理论与方法建立了老年道路交通抉择行为模式，从根本上阐明了个人、家庭经济状况等社会背景因素对老年道路交通抉择行为的影响，从而预测了对老年道路交通抉择行为变化规律[36]。刘炳恩基于2003年的北京城市住户交通问卷调查数据分析，探讨了对城镇居民交通方式选取的主要影响因素，选择了实际日常生活中较为普遍的出游办法作为对城镇居民交通办法的重要选取，包括公交、出租车、私家车等共计五种方法，并确定了对城镇居民交通选择方式发生影响的特性变量以及其对应的计算办法，从而构建了城市出游方法的选择MNL模型[36]。何明等在合理利用RP和SP调查数据的条件下，成功估计了其他一种出游方法游客向都市轨迹交通出行方式的转化率，并建立了适合于预计都市轨迹交通出行方式的非集计MNL模型，以及用来预计苏州都市轨迹出游方法客流量的实际案例研究[36]。刘崭等将城市轨道公共交通与地面常规公共交通作为共同选择肢，并联系西安市城乡居民公共交通出游特点，确定交通影响因子的个人特征、国家特点等共同出游特征，形成了城乡居民公共交通出行方式选择的MNL模式。李世民等在非集计模型的条件下，对模型中的时间作用参数加以推理，以获得求解时间价值问题的有效工具[36]。通过对乌鲁木齐的市民交通调研数据对效用函数开展了标定分析，可以确定产生影响出行时间价格变化的各种因素，同时也产生了在不同地区一样家庭收入、旅行目的、出游时段等的人对交通方式做出抉择时出行时间价格的变化情况[36]。宗刚等人基于2010年对北京城市住户交通问题调研数据分析，建立了基于家庭决策的北京住户城市交通模式，选取非集计模型[36]。结果表明，从家庭和个人这二种不同出行方式链层次上，模型都能够很好地刻画上班出行者的交通方式选取情况。顾宇等人利用2009年的南京市民城市交通问卷调查数据分析，运用非集计模型定义，选取了公交和非机动车对上班出游的若干影响要素，并计算和解析出不同因素对选取结果所产生的负面影响，同时运用SPSS软件建立以出行为单位的公交和非机动车出行方式选择的BL模型，并解析各因素影响程度[36]。结果显示工作出行中的主要影响因素，从高至低依次是性别、持有非机动车量、对私家小汽车的看法以及旅行耗时等，同时上述变量对选择公共交通或者非机动车出行的影响也更加突出。综上所述，出行者对出行方式的选用与出行者的个人特点和旅行性质、外部环境等因素有关。非集计模型还能很好地检验上述变数，为城市交通分析问题提供了更加科学而有效的解决方案。由于非集计模型分析的调查单位都是真正产生实际交通管理活动结果的单位，因此应该将不能作为任何数据分析基础的调查资料，直接拿去用以建立模型[36]。与以往的集计模型作对比，非集计模型具有以下突出优点[36]：(1)以确定的社会行为假设当作基准，具有更加科学的逻辑性。(2)模型的自变量都可以用相当少的数据标记出。(3)自变数都可以使用和个人选择相关的元素，有利于更为准确地描述个人或居民的交通决策流程。(4)模型具有良好的区域变化性和时段变化性。(5)用于衡量各项道路交通管理计划和交通管理措施的有效性。

接下来是关于结构方程模型在出行选择方面应用的研究现状。结构方程模型的广泛使用国外人们对于将构方程的使用已相对地较为成熟了，但在对于居民日常生活行为的科学研究领域，结构方程模型也越来越变成了一个主要的科学研究方式。Lu和Pas等人把对个人、家庭的社会经济因素，以及社会活动方式视为影响旅游行为的最主要变量，并使用了社区结构方程模型,以探究通勤者活动参与、社交因素和旅游行为之间的关系关联[40]。Golob通过对波特兰区域市民的出游活动开展研究，将通勤活动作为重点研究内容，利用结构方程模型研究出游链数量、活动时间、出游时间三种相关因子间的交互关联。Golob以家为主要研究对象，并以家庭空间的可达度作为主要外生要素，通过运用空间结构方程建模方法，将一家人出游时间、旅行量的产生过程模式化,进而探究了一家人出游的时间活动量与家庭成员数量之间的联系关系[[40]。ChristianA，Kckner从个人出游意愿,社会行为规范,及外部环境和行为等的角度剖析了环保友好型出游行为[40]。其中,最重要的假说来自计划行为理论(TPB)和规范激活模式(NAM),而上述各种因素最后都被整合到了一种模式中,称之为CADM模式[40]。并运用结构方程模型针对三百八十九个学生的数据进行了检验，结果认为主体和客观环境变量会对大多数的旅行方案选择影响，不过影响最大的是个人旅行习惯和出游意愿。

目前在国外的科学研究中，使用结构方程来研究在旅行过程中会影响到旅行方案所选取的不可测量原因，不过目前在国内的实际应用还比较少。国内外的研究者也逐渐开始使用这些方式来研究我国各城市的居民出行方式选择。张文佳等人就从家庭的视角分析了城镇居民对出游的需要，从而构建了中国城镇居民出游需要模式，并利用结构方程模型进行了证明[40]。李霞等人同样利用结构方程建模技术，对节假日期间市民出游的活动进行了模型解析，并探讨了市民出游行为和活动之间的关联[40]。杨励雅教授认为，城镇居民出行方式的选择主要受个人和家庭的社会经济属性、出游链条、土地使用三类因素的共同影响，其中个人和家庭经济因素是外生变量，而出游链条、土地使用因素及其出行方式的选择则属于内生变量，并从出游链条、土地使用因素等方面，系统阐述了城镇居民出行方式选择的主要影响因素，对于城镇居民出行结构的综合优化具有重要的指导[40]。陈刚等人在对停车换乘服务价值进行评估的过程中，引进了公众参与性理论，并研究了城市外围小汽车出行者对停车换乘的特征及其对换乘设施的服务，从而形成了小汽车出行者对其换乘服务价值评估的结构方程模式，即小汽车出行者可以通过行为的特征与行为的特点判断对换乘设施服务的要求，进而通过对比服务要求和换乘服务本身的服务， 从而做出价值判断。龚韵枝教授在运用SP研究结果后，根据都市外围出游者的个人特点、经济活动特性和方式的选取特点，运用结构方程模型，真正对都市外层住户之间停车位换乘的方式选取特点做出了模型，以一种崭新的视角对都市外层停车位换乘设施需求量做出了预估[40]。刘武等人运用结构方程创建了中国城市公交服务质量旅客满意指标模式，提出了满意度与旅客期望、乘务员满意度、感知质量、乘务员忠诚度等因素间存在关系，并运用均值的比较分析方法，为文中选取的二十五条线路实行了排名分类，得出了相应的满意指标[40]。周钱等人通过结构方程模型分析和模拟出行者的特性、社会活动参与和交通活动之间的相互作用关系，并用以研究城市交通需求，从而证实了构造方程模型在研究城市交通需求过程中的有效性[40]。曹小曙等人通过对广州的三个居民小区进行调查研究，通过构建居住与出游结构方程模型分析了居住属性、外出时机、居住地位，和住户出游目的、出游方法、出行偏好间的关联，并从性别、年龄段、收入和居住地位等视角，解析了住户出游决策与形成的基本机制[40]。

* 1. 文献述评

从上述三大方面的文献综述来看，一些地铁新技术在居民出行过程中以及常态化，比如二维码刷码过闸已经成为了各城市地区轨道交通行业首选的支付和过闸方法，而地铁作为城市内部主要的大型轨道公共交通工具是非常具有代表性的。同时根据居民出行选择影响因素的研究可得，地铁新技术属于影响居民出行选择的制度技术情境因素，并且这将会影响到居民的出行习惯，环保意识以及态度等多重因素。同时，居民出行选择将会对环保节能产生直接影响。根据现状可知，现已有很多公交车和出租车对居民出行选择影响的研究，而在轨道交通与居民出行选择关系方面的研究却非常少。因此研究地铁新技术发展与居民出行选择之间的关系是非常有意义的，同时能够对未来的绿色环保发展产生积极影响。

同时，根据非集计模型在居民出行选择方面的研究现状可得，建立Logistic回归模型和结构方程模型通常是联系居民出行选择和其他研究因素的首选方法，因此本文也将利用Logistic回归模型对居民出行数据进行探究，并用结构方程模型进行验证，以此来寻找居民是否选择乘坐地铁出行与各变量之间的相关性。除此之外，本文还将基于大数据对客流量变化、时间差变化进行统计分析，还将对数据进行空间模式挖掘来探究居民出行选择与地铁新技术政策实施的关系。

1. 基于大数据的地铁乘客行为模式挖掘
   1. 数据处理
      1. 数据获取及说明

由于目前与地铁新技术相关的数据较少、每个城市对于政策的相应时间存在差异且地铁的使用还未具有普遍性，因此在考虑了数据获取的难易程度、政策实施及响应程度和对地铁使用的熟悉程度以后，我们优先选择了北京市的地铁乘客出行数据作为研究数据来探究地铁新技术与居民出行选择之间的关系。

二维码刷码过闸这项技术在北京市正式投入使用的时间为2018年中旬左右，因此我们选择政策颁布前的数据和政策颁布一段时间以后的数据作为对照组来对地铁乘客的前后行为变化进行研究。数据的主要内容有出站时间、到达车站编号、卡号、其实车站编号、进站时间、卡大类、卡的子类型、扣费金额、对账日期、结算日期、安全密钥、闸机号、运营单位、钱包余额以及交易序列号。一共在高德以及地铁官网获得了1亿6000万条数据，通过数据洗涤和清理最后我们将得到8000万条零散的可用数据。

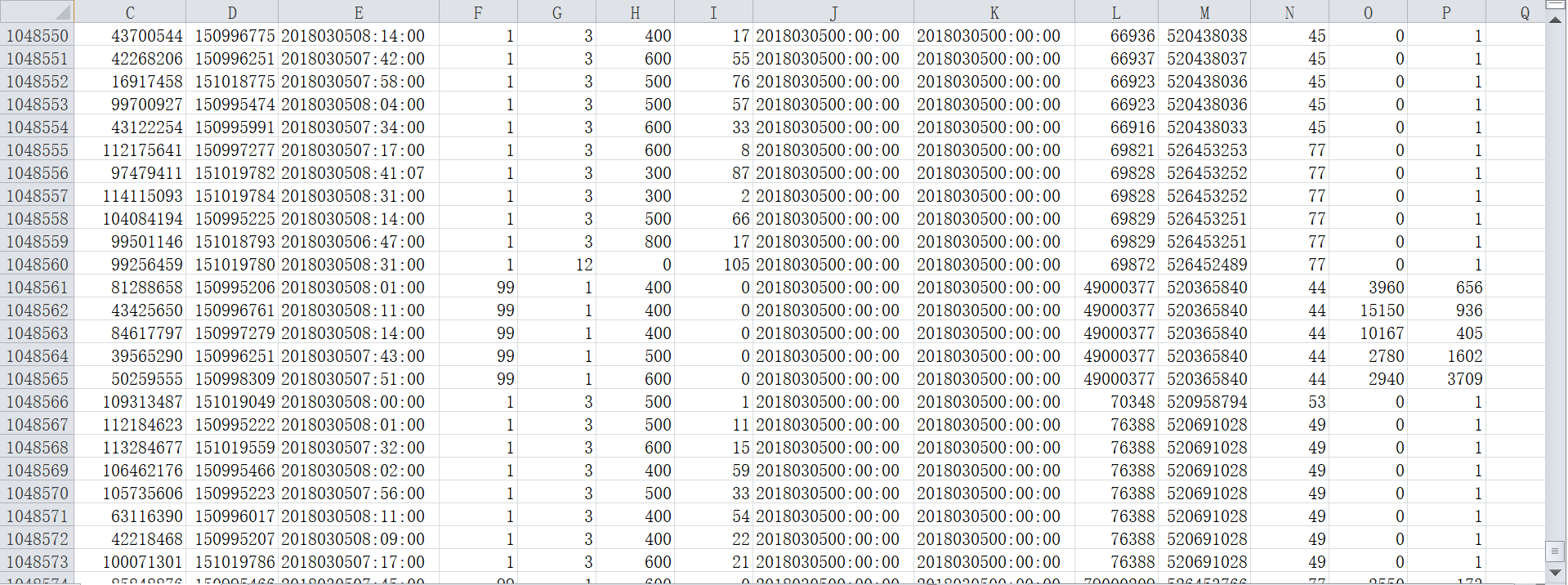


图3-1 2018年部分数据示例图

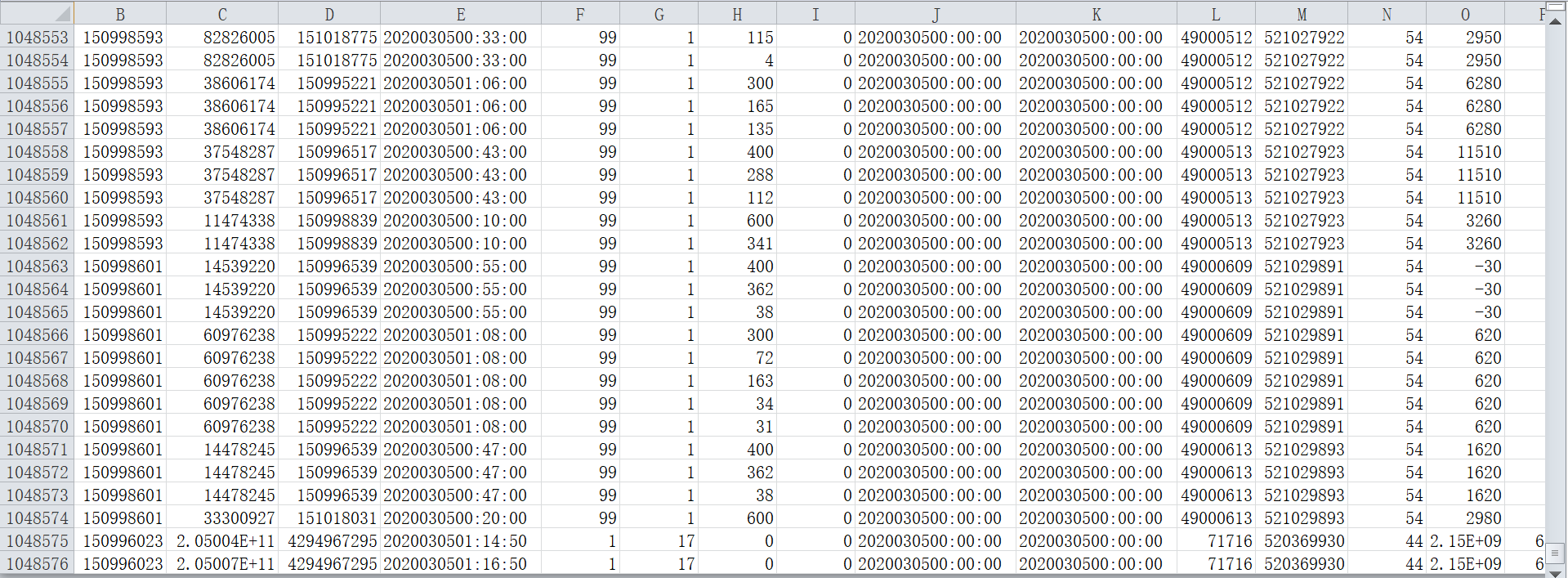


图3-2 2020年部分数据示例图

* + 1. 基于Python的数据整合

为了之后关于前后两年的居民出行时空模式聚类分析的研究，我们需要将8000万条数据按年份进行整合。由于数据量过于庞大，因此我们只能选择python作为工具来对数据整合。

首先调取数据集，并利用python中的append语句和sort语句对数据进行整合排列，具体过程如图3-3。



图3-3 python读取文件路径图

下一步，设计表的存储格式、路径、title以及每一行数据的顺序排列，具体过程如图3-4。



图3-4 python创建表格图

具体运行过程如图3-5。



图3-5 2018年数据运行过程示例图

最后运行完，2018年和2020年分别都整合了4000万条数据，共计8000万条数据这些数据将用于后期的时间模式挖掘以及空间模式挖掘之中。

* 1. 基于客流量统计的乘客行为分析

已知北京市二维码刷码过闸投入使用的政策颁布时间为2018年下旬，因此我们可以先在政策提出前后各自随机选取6个月的居民出行数据，然后统计不同年份相同6个月数据中的到达车站编号出现的频率，到达车站编号出现的频率在这里可以近似等于车站的客流量。为了直观的看出变化且避免选到某些受到极端影响的数据，我们将会选取出一些频率值比较稳定且数值较大的到达车站编号，观察政策提出前后这些到达车站编号的频数变化即客流量的变化来初步确定该政策的提出是否会对居民出行选择产生一定的影响。由于数据量过于庞大，已超出SPSS所能承载的最大数值量，因此我们将会每次只向SPSS中导入一个月的数据进行频率统计，然后再将统计出来的到达车站编号6个月的频率值进行加和，具体过程如下。

以2018年和2020年1月份为例，将数据导入到SPSS中，然后进行分析—描述统计—频率。

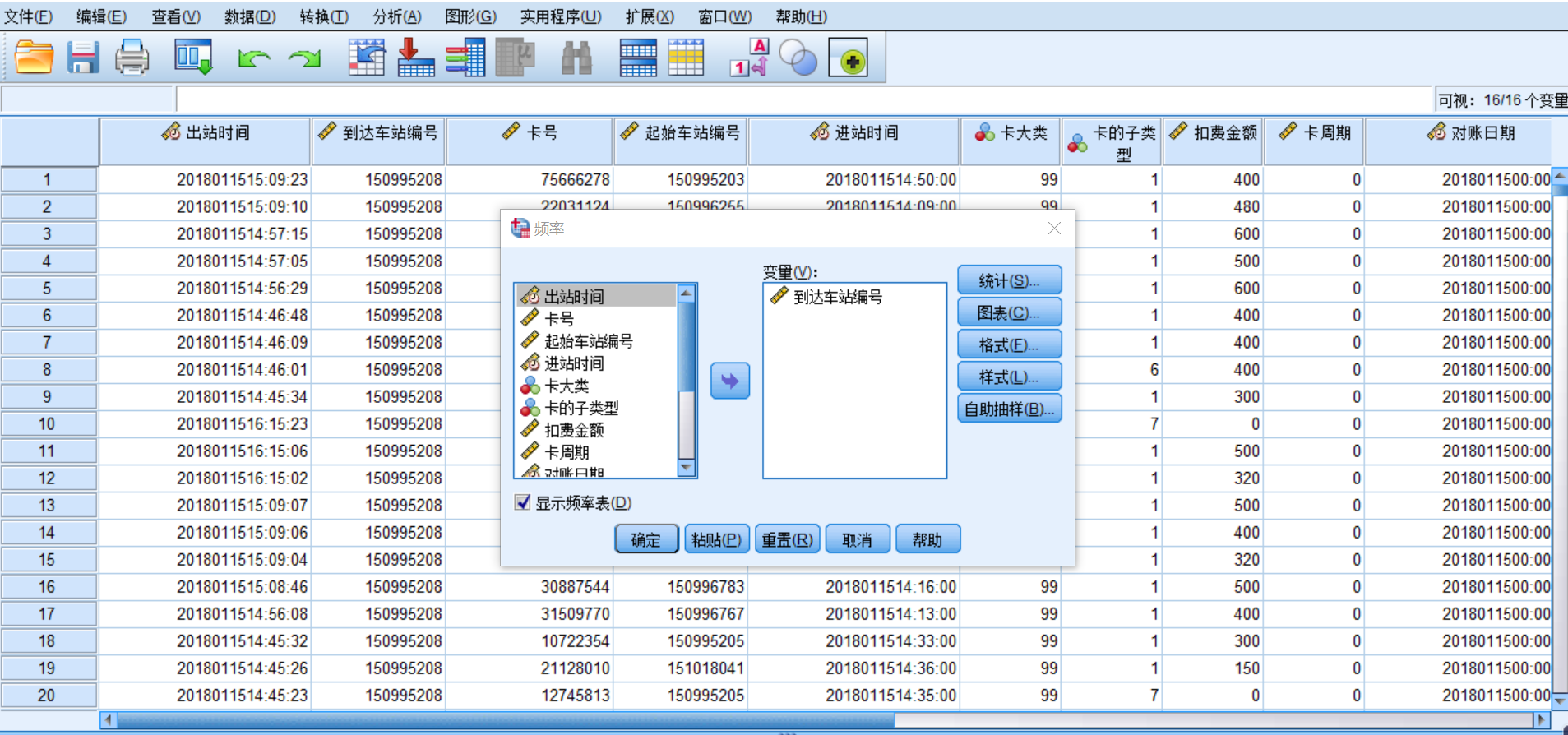


图3-6 2018年数据SPSS运行图

得出以下频率图表。

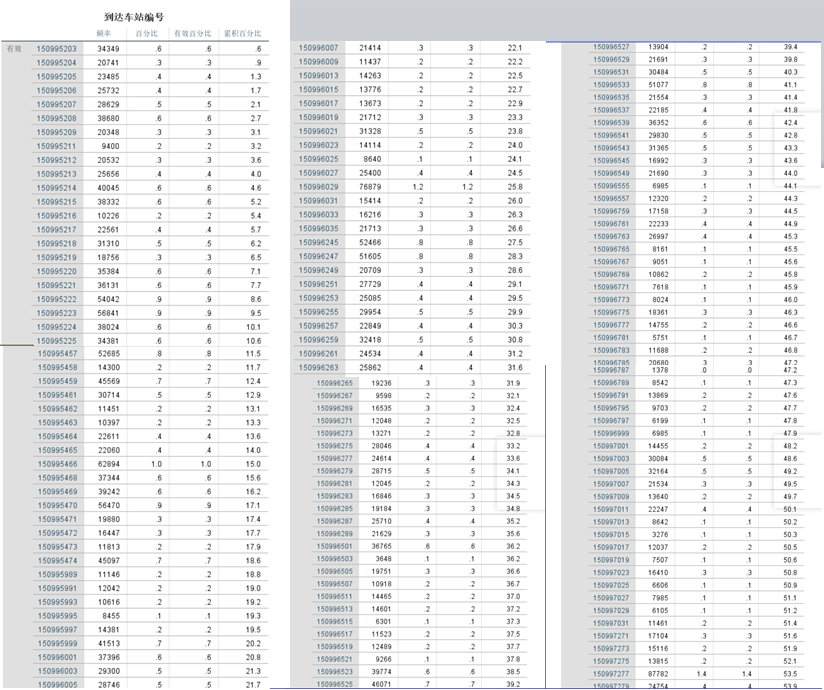


图3-7 2018年部分数据频率图示例1

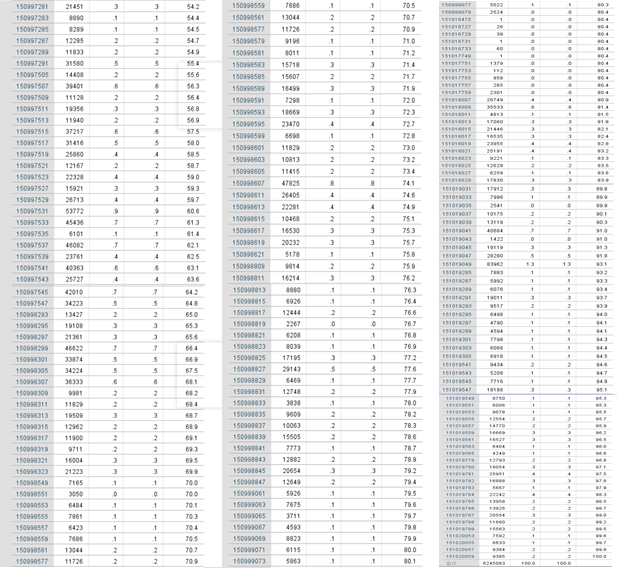


图3-8 2018年部分数据频率图示例2

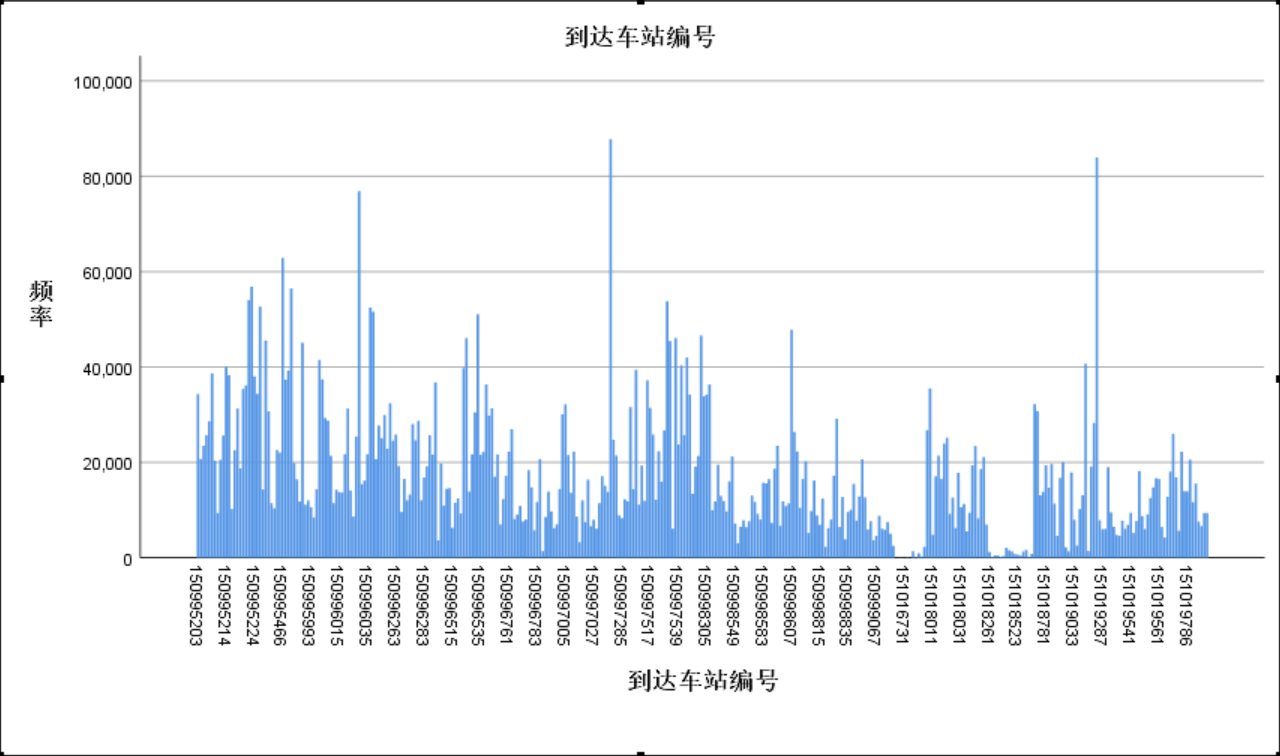


图3-9 2018年到达车站编号频率图

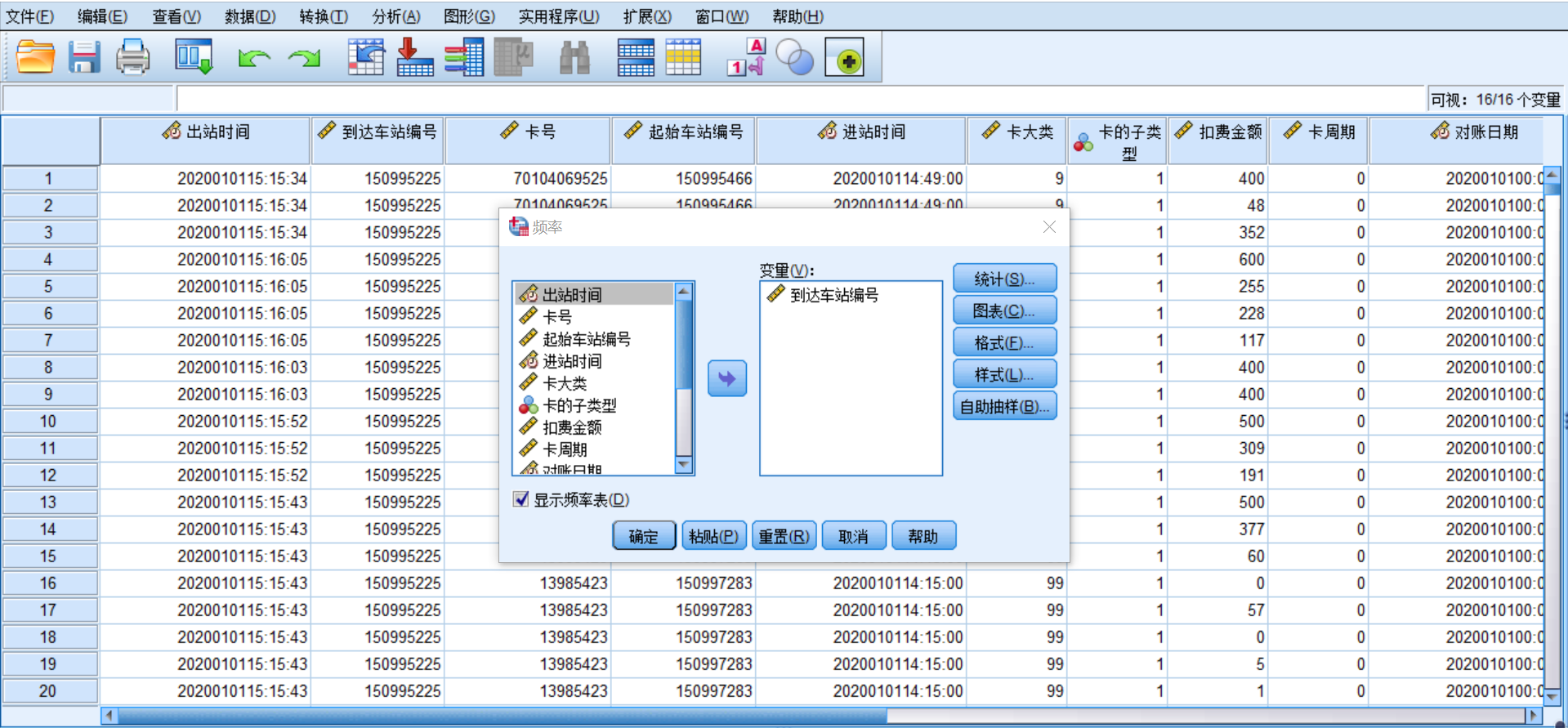


图3-10 2020年数据SPSS运行图

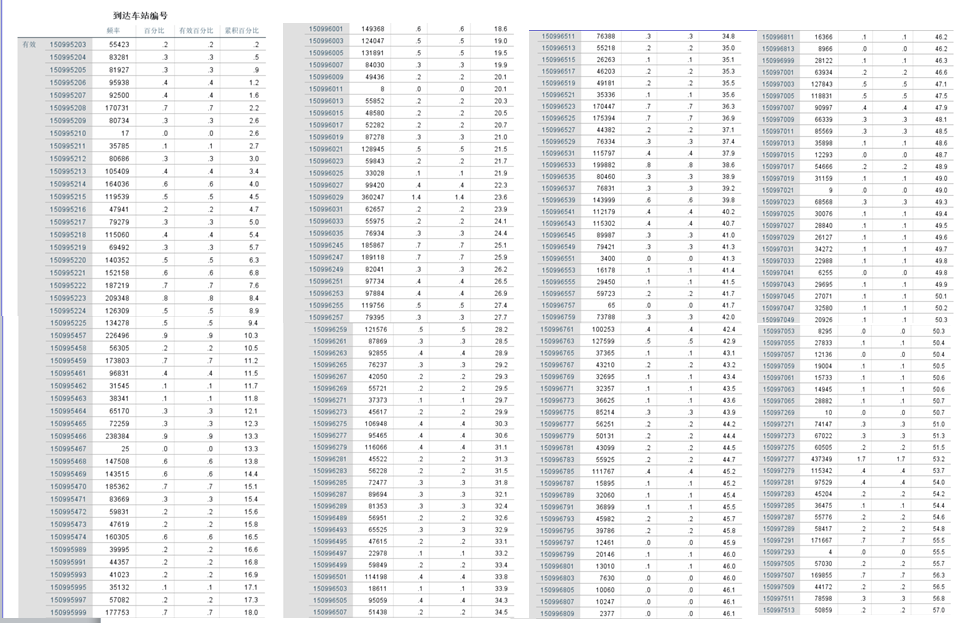


图3-11 2020年部分数据频率图示例1

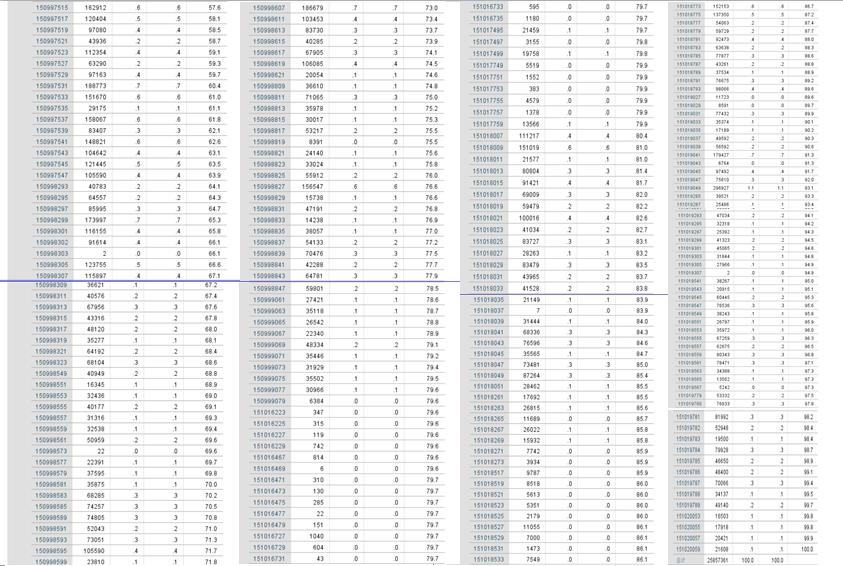


图3-12 2020年部分数据频率图示例2

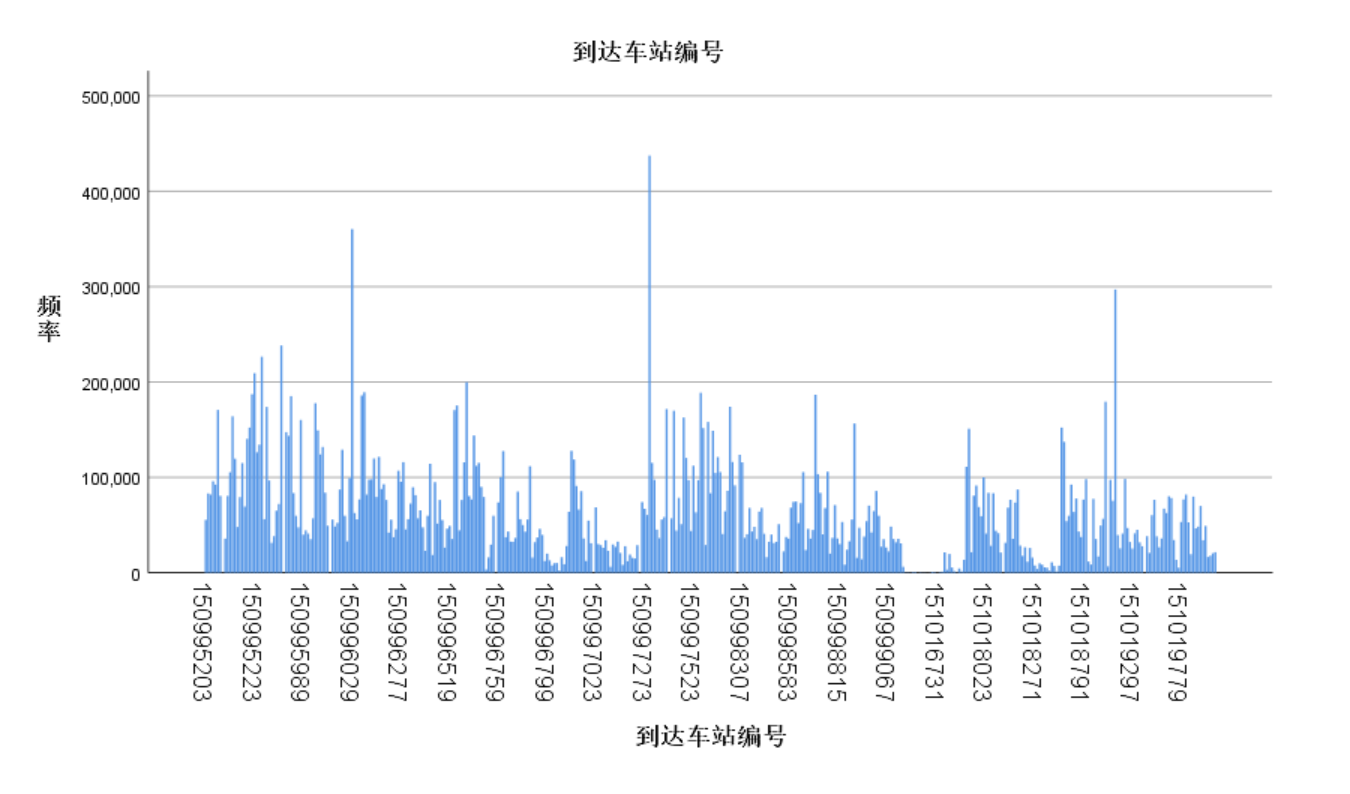


图3-13 2020年到达车站编号频率图

如图可以看出150995203，150995224，150996029，150990533，150997282等到达车站编号频数更高，因此我们将选择这一批到达车站编号进行分析。与上述过程相同，最终我们共选取了350个车站编号，将350个车站6个月的频率相加做和，最后统计出2018年6个月的客流量大约有42483000人，2020年6个月的客流量大约有42496792人。我们可以看出在该政策提出后，2020年的客流量比2018年客流量增长了0.03%，几乎没有发生变化，根据实际情况我们得知2020年的客流量很大程度上受到了疫情影响，所以研究结果还需要进一步考究。

* 1. 基于时间差变化的乘客行为分析
     1. 时间差指标及获取

目前拥有的数据变量中包含地铁乘客进出站的具体时间，因此我们可以将进出站时间差的变化作为在时间维度上衡量地铁新技术的使用对居民出行选择影响的指标。如果20年的平均时间差远远小于18年的平均时间差，那么证明二维码刷码过闸能够大幅度的降低居民进出站的时间，提高了地铁乘客出行效率，从而使更多居民选择乘坐地铁出行。因此，这就需要整合18年和20年的居民出行数据。由于数据量过于庞大，所以需要利用python对8000万条数据进行时间差计算并导出。具体运行过程如图3-14。

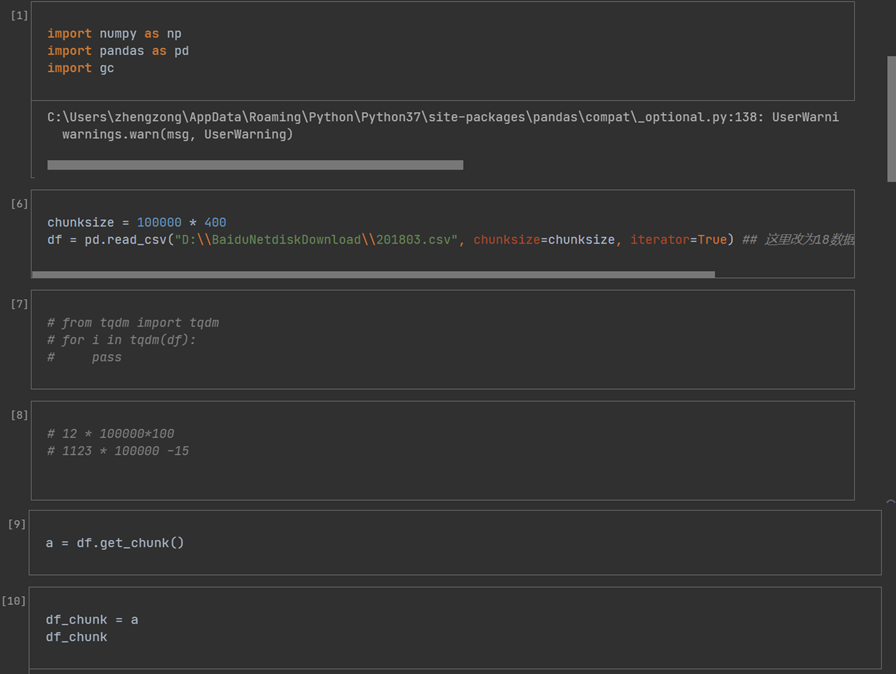


图3-14 2018年整合数据读取过程图

读取的运行结果表如图3-15。

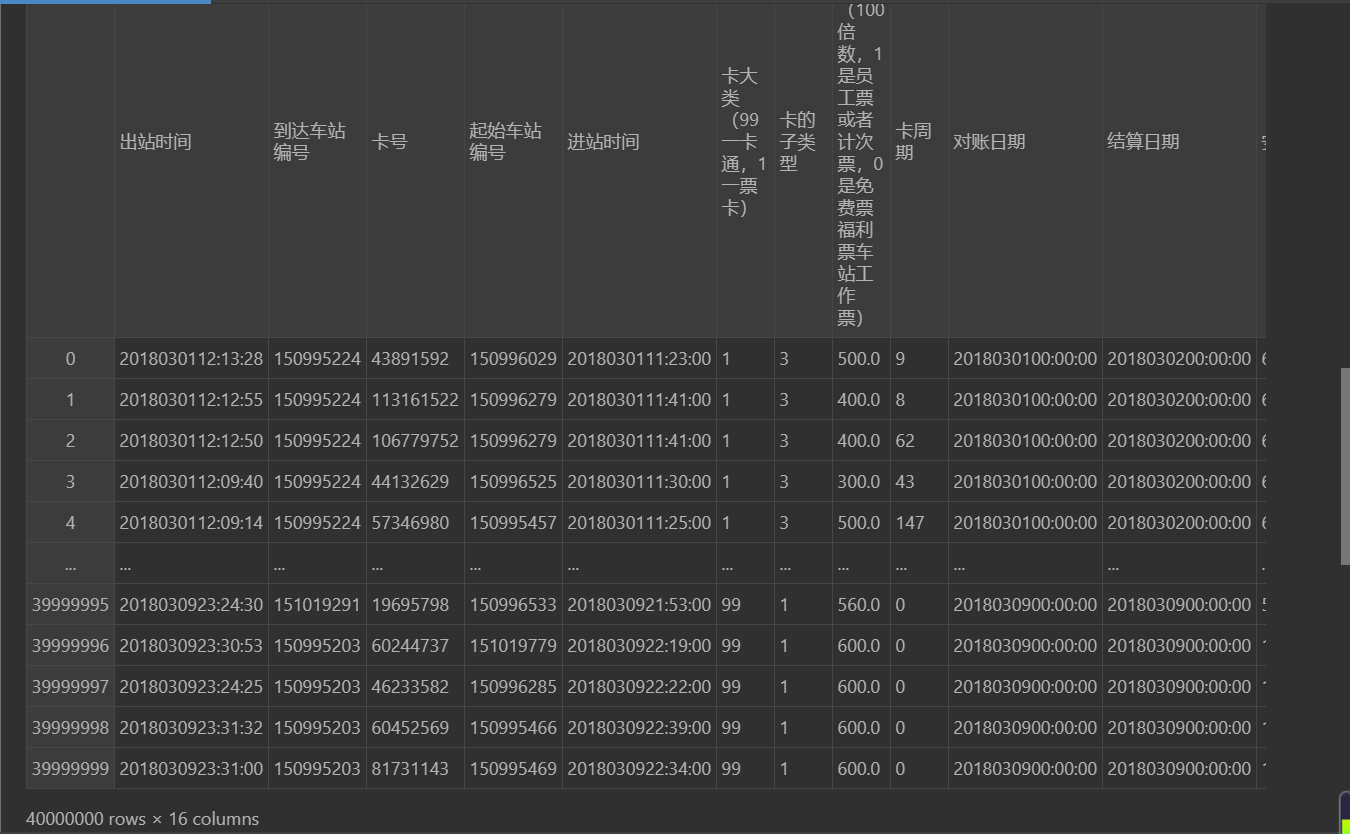


图3-15 2018年整合数据运行结果图

将进站时间和出站时间转化为时间格式，并用出站时间减去进站时间求出时间差，最后将数据集进行保存，具体过程如图3-16。



图3-16 2018年时间差计算过程图

最后将数据导出至excel表格，用于后期的时间模式挖掘和空间模式挖掘，如图3-17。

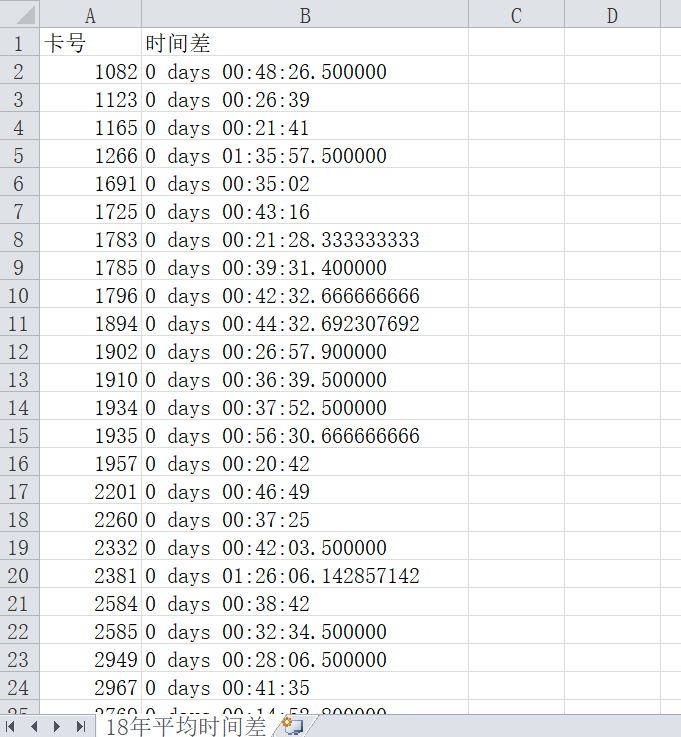


图3-17 2018年时间差数据图

2020年的时间差数据获取方式与2018年相同，见附录。

* + 1. 基于时间差变化的统计分析

在上一节中我们已经分别得到了2018年和2020年的时间差数据。由于数据量过大，并且其中还存在一个卡号可能只存在于一年的现象，因此我们无法直接将两年的时间差进行比较得出结论。所以，我们需要对两年都存在的卡号进行提取，然后利用时间差变化指标对乘客出行选择行为的变化进行时间维度上的模式挖掘，这里主要利用模式挖掘中的聚类分析法来探究两年的时间差变化。如果20年的时间差比18年的时间差增加，则说明二维码刷码过闸的投入使用并没有对居民出行选择产生积极影响。反之，如果20年的时间差比18年的时间差减小，则证明二维码刷码过闸技术的使用对居民出行选择产生了积极影响，具体过程如下。

首先设置前后两年的时间差变化的计算过程，如图3-18。

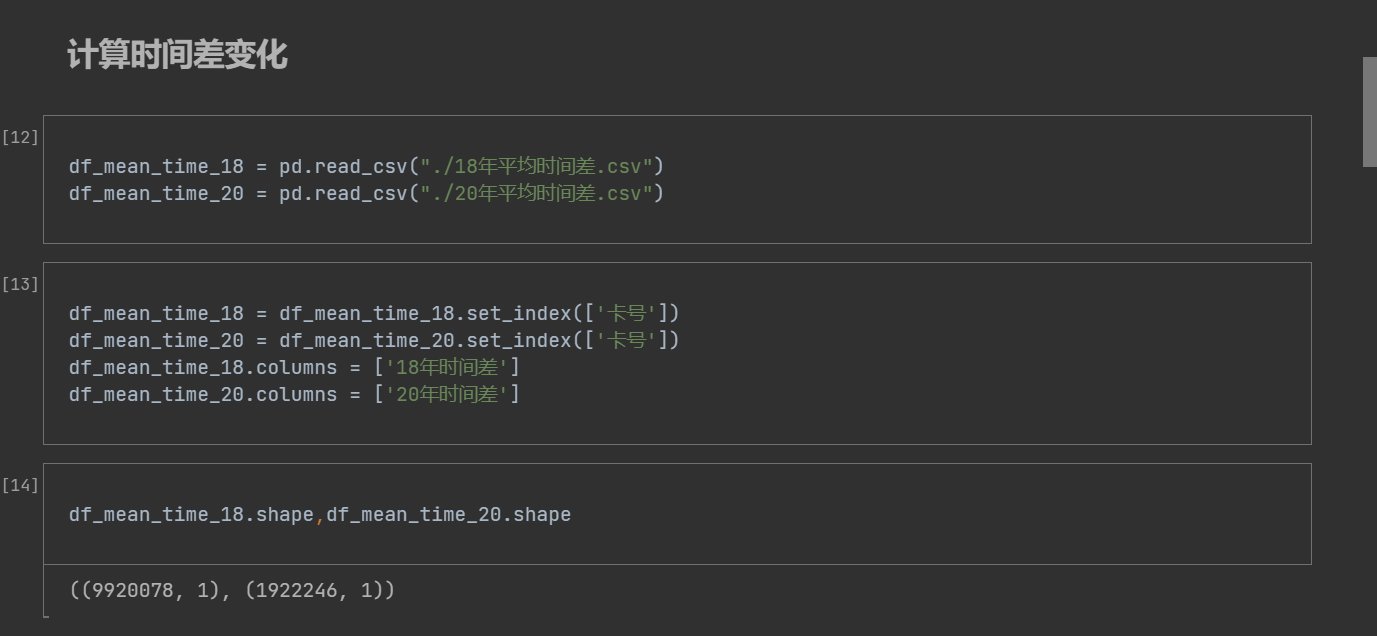
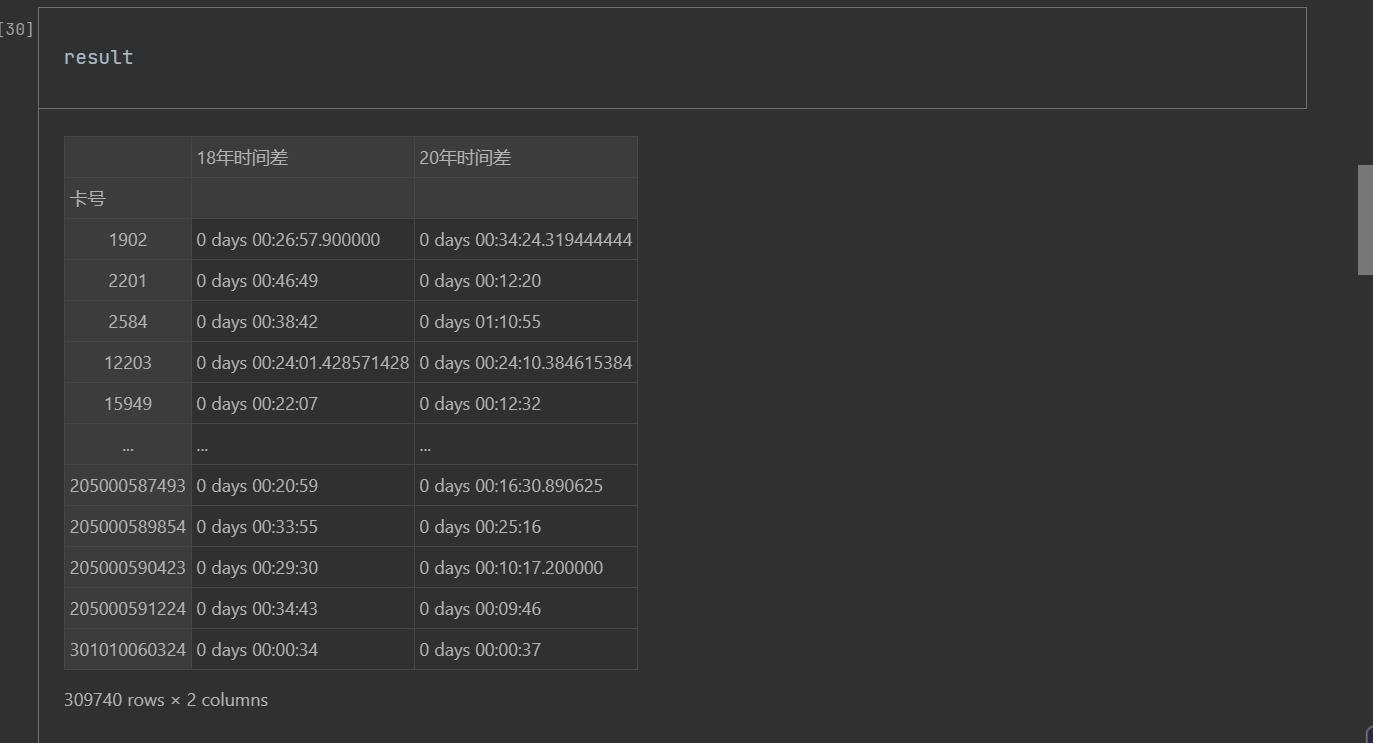


图3-18 两年时间差变化计算图

然后读取18，20年时间差数据中都存在的卡号，最后得到了309740条两年都存在的不同卡号的时间差数据，如图3-19。

图3-19 时间差数据图

用20年的时间差减去18年的时间差，并将时间差换算成小时和分钟为单位，如图3-20所示。

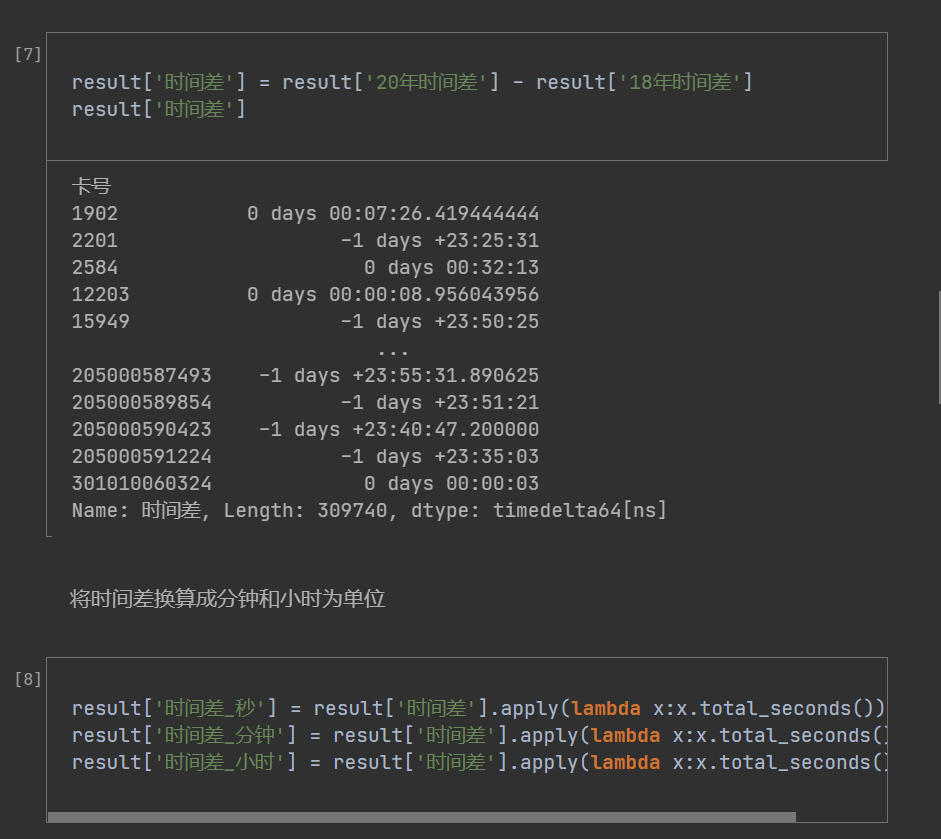


图3-20 两年时间差变化运行图

得出以小时和分钟换算的两年时间差的表格，如图3-21所示。

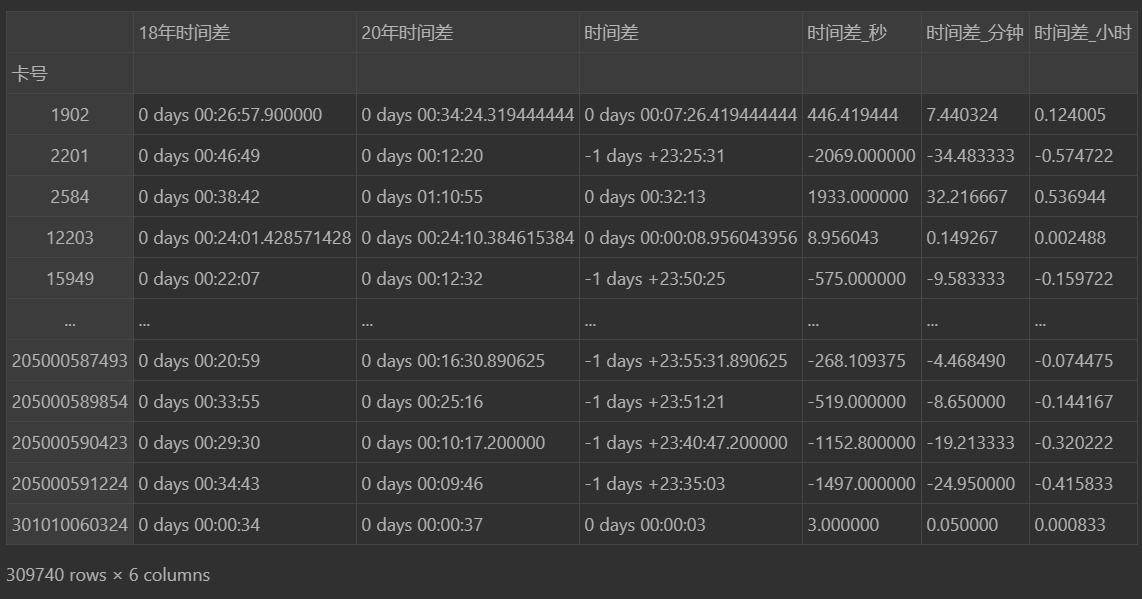


图3-21 两年时间差变化结果图

接下来，对表格里的数据建立以分钟为单位的密度直方图来观察时间差的集中分布情况，如图3-22。

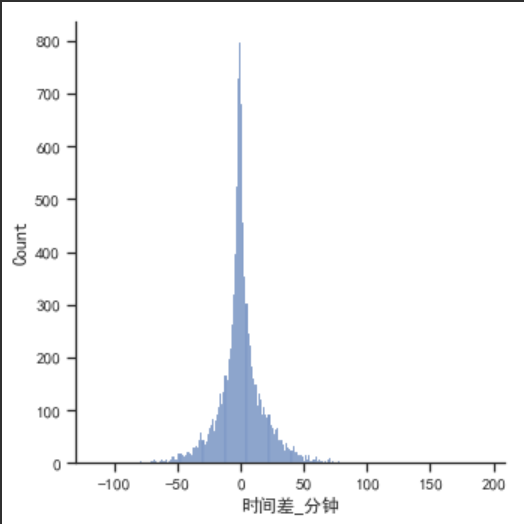


图3-22 两年时间差变化条形图（以分钟为单位）

通过上图可以看出，两年的时间差基本呈现以0为均值的正态分布，基本集中在-50至+50分钟的变化之间。然后再对表格里的数据建立以小时为单位的密度直方图来观察时间差的集中分布情况，如图3-23。

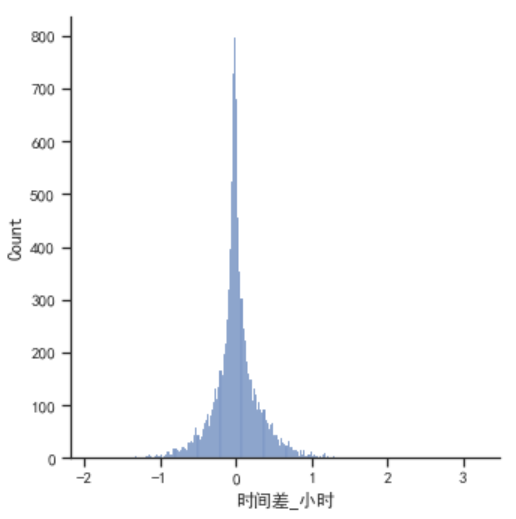


图3-23 两年时间差变化条形图（以小时为单位）

同理，我们可以观察出，两年的时间差基本呈现以0为均值的正态分布，基本集中在-1小时至+1小时的变化之间。

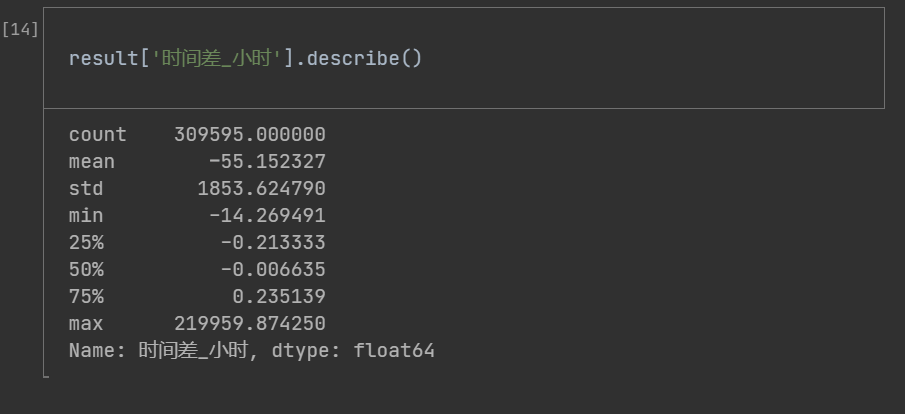


图3-24 两年时间差变化计算结果图

最后通过数据比较可以明显的看出，做差后的均值是-55.152327小时，也就是说20年的时间差在总体情况上是要小于18年的时间差的，即二维码过闸技术的使用能够缩短居民进出站时间，从而可能会使更多的居民选择乘坐地铁出行。

* 1. 基于空间模式挖掘的乘客行为分析
     1. 活跃乘客及活跃天数计算

目前拥有的数据变量中包含了地铁乘客的卡号以及钱包余额，地铁乘客卡号出现的次数可以近似于乘客的活跃天数，因此我们可以通过聚类分析的方法来对乘客活跃人数和钱包余额建立联系，从而挖掘出20年相比于18年的活跃人数变化。如果20年的活跃人群数量大于18年的，则证明该政策的实施对居民选择乘坐地铁出行产生了积极影响。如果20年的活跃人群数量小于18年的，则证明该政策不起作用或对居民选择地铁出行产生了抑制作用。因此为研究这一问题，我们首先需要分别对18年和20年每一个卡号的活跃天数和钱包余额进行统计整合，然后再进行下一步的聚类分析。由于数据量过于庞大，所以需要利用python对8000万条数据进行活跃天数和钱包余额的整合统计并导出。

首先调取出18年的总整合列表，然后再以每个卡号的不同日期出现次数为基准，对总列表中的每一个卡号出现的频率以及其钱包余额进行统计，具体过程如下图。

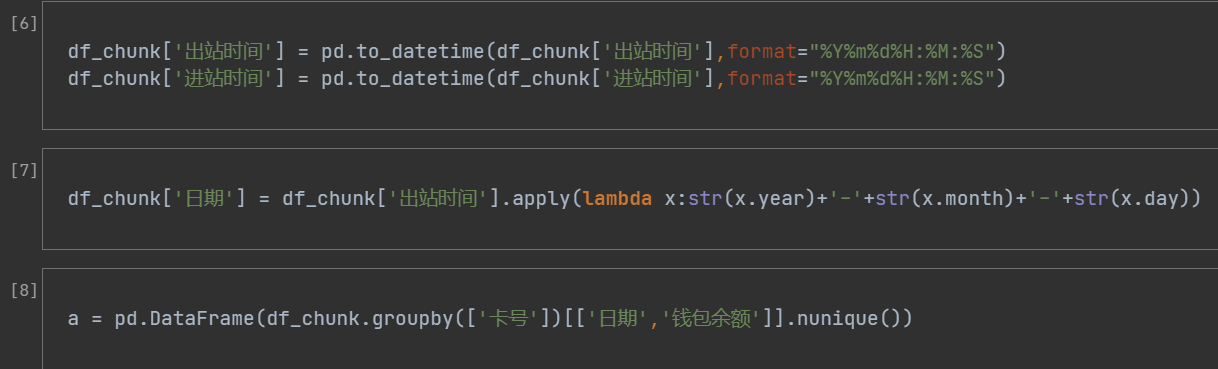


图3-25 调取整合列表[44]

运行结果如图。

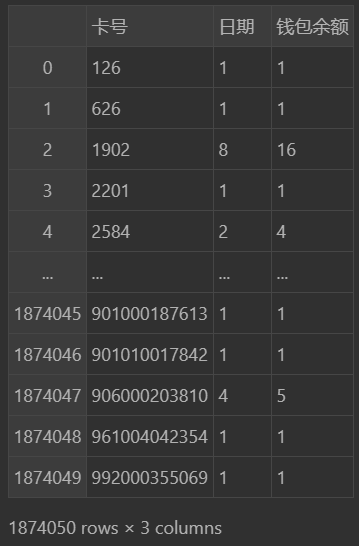


图3-26 2018年不同卡号日期及余额统计图

最终一共得到了1874050条不同卡号的频数统计数据以及其钱包余额总数统计数据。

再调取出20年的总整合列表，然后再以每个卡号的不同日期出现次数为基准，对总列表中的每一个卡号出现的频率以及其钱包余额进行统计，整合结果如下图。

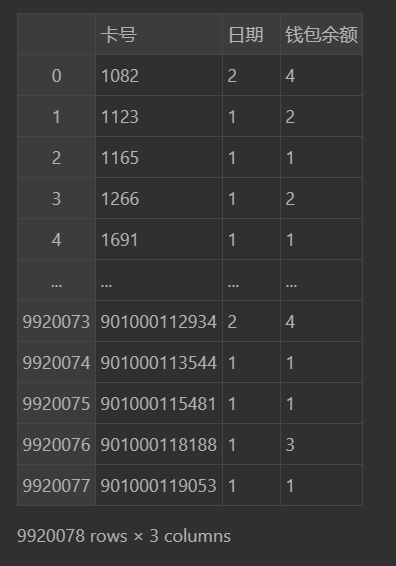


图3-27 2020年不同卡号日期及余额统计图

最终一共得到了9920078条不同卡号的频数统计数据以及其钱包余额总数统计数据。

* + 1. 基于活跃乘客及天数的聚类分析

根据上述的整合数据，对2018年和2020年的卡号、活跃天数以及钱包余额变化进行聚类分析，设置类别1为活跃人群，用蓝色来表示，类别0为非活跃人群，用红色表示，具体过程如图。



图3-28 钱包余额及活跃天数的聚类过程[44]

2018年和2020年的聚类结果分别如图所示。

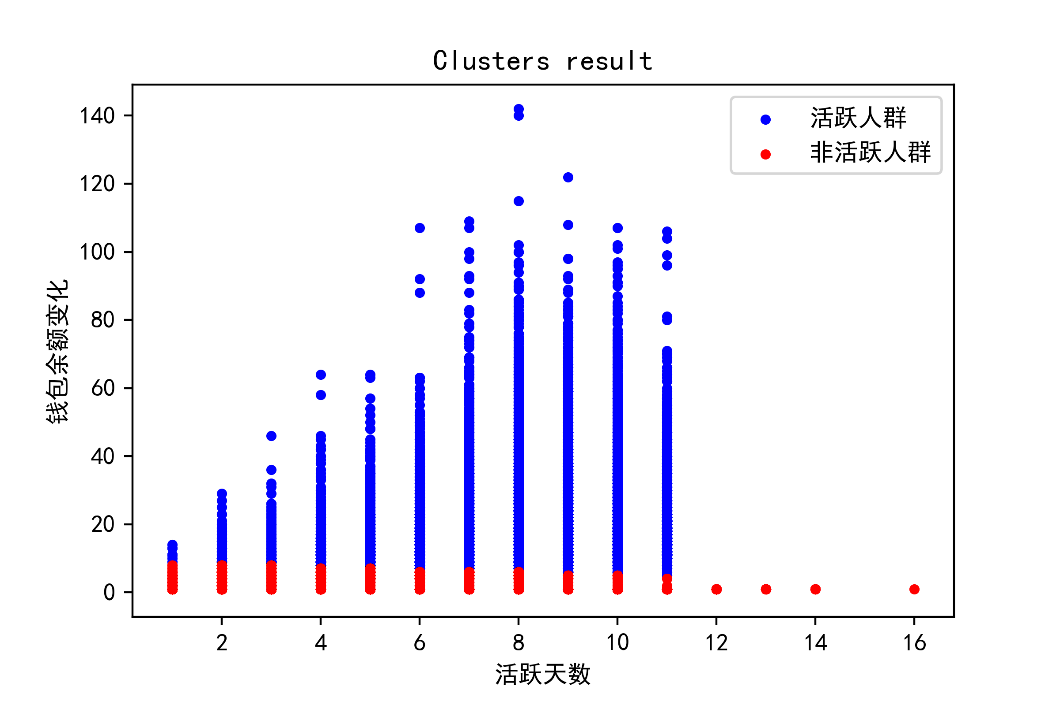


图3-29 2018年活跃人群聚类结果图

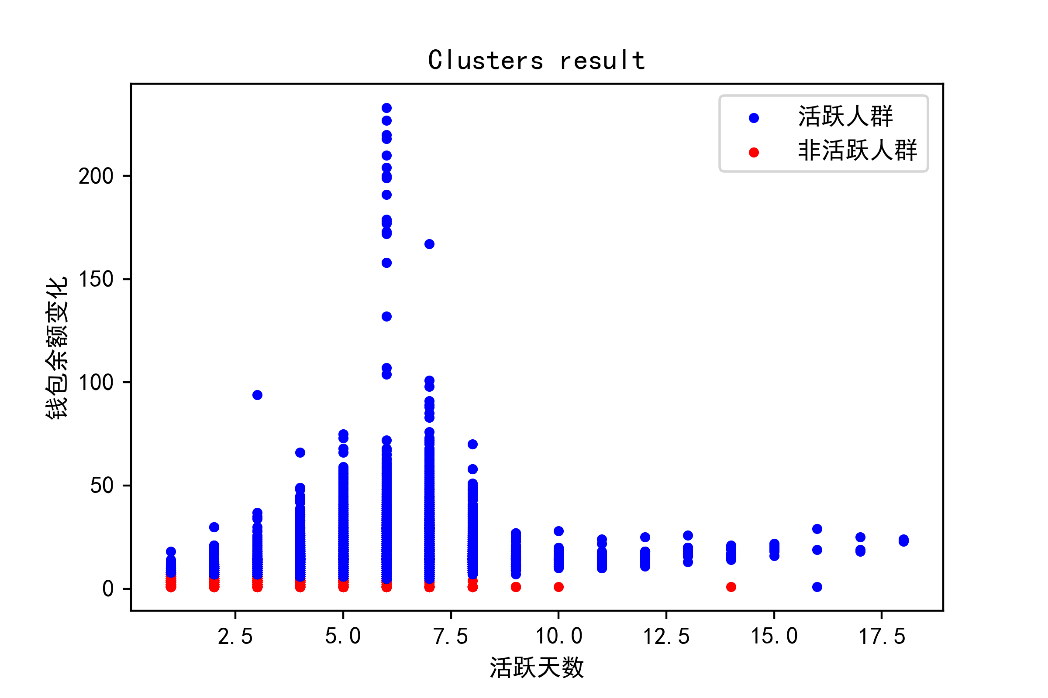
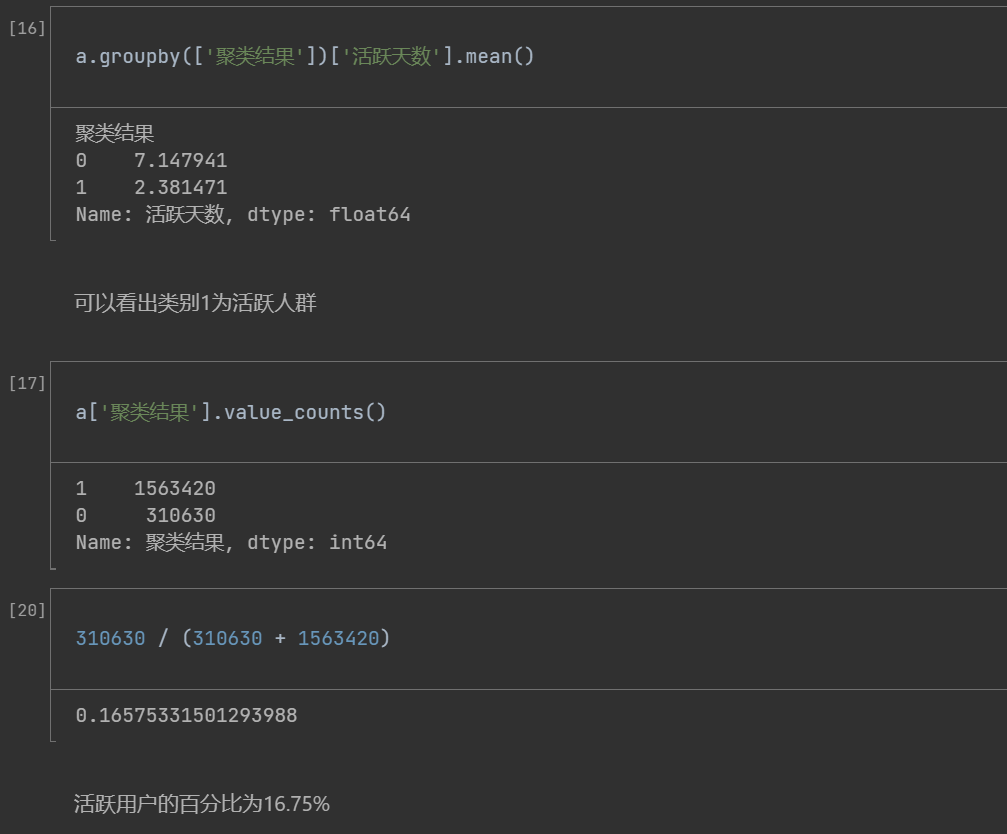


图3-30 2020年活跃人群聚类结果图

已知蓝色表示活跃人群，红色表示非活跃人群，建立的是钱包余额变化和活跃天数之间的聚类关系。从上图聚类结果可以看出，虽然蓝色远远多于红色，但这并不是代表活跃人群的人数远大于非活跃人群的人数，而是表示活跃人群钱包余额变化幅度要更大，通过观察可以看出非活跃人群的钱包余额变化次数大部分都少于10次，而活跃人群的钱包余额变化要更加频繁。从18年图与20年图的对比，可以发现活跃人群的活跃天数明显增加了。在18年的图中，活跃人群的活跃天数基本全部小于12天。而在20年的图中，可以看到很多活跃人群的活跃天数分布在10-17.5天的区间内。

最后再计算2018年和2020年活跃用户百分比，如图所示。

图3-31 2018年活跃用户百分比[44]

根据聚类计算，2018年非活跃人群共有156342人，活跃人群共有310630人，活跃人群占总人数的比例为310630 / (310630+1563420) = 16.75%。

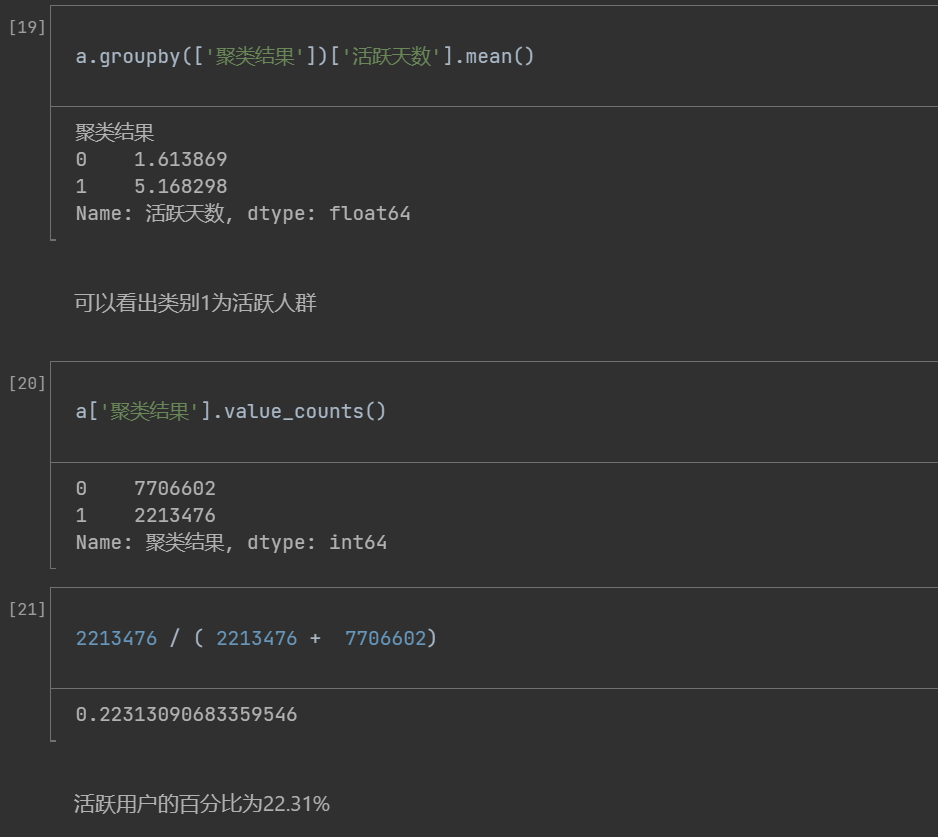


图3-32 2020年活跃用户百分比[44]

根据聚类计算，2020年非活跃人群共有7706602人，活跃人群共有2213476人，活跃人群占总人数的比例为2213476 / (7706602+2213476) = 22.31%。

从聚类结果可以看出，从18年到20年，活跃人群的比例从16.75%增加到了22.31%。

* 1. 研究结果分析

该部分主要通过从统计分析和模式挖掘两个角度来研究地铁新技术政策的实施与居民出行选择之间的关系。通过客流量变化的统计分析可得，2020年的客流量变化比2018年的客流量变化增加了0.03%，但是客流量变化受疫情影响较大，且变化量较小。通过时间差变化可得2020年的时间差减去2018年的时间差的均值是-55.152327小时，因此20年的时间差总体上小于18年的时间差，可以证明二维码过闸技术的使用能够缩短居民进出站时间，从而可能会使更多的居民选择乘坐地铁出行。通过对活跃天数和钱包余额进行聚类分析可以得到20年相比于18年活跃人群的变化结果，从18年图与20年图的对比，可以发现活跃人群的活跃天数明显增加了。在18年的图中，活跃人群的活跃天数基本全部小于12天。而在20年的图中，可以看到很多活跃人群的活跃天数分布在10-17.5天的区间内。

再根据最后的聚类计算可知，2018年非活跃人群共有156342人，活跃人群共有310630人，活跃人群占总人数的比例为310630 / (310630+1563420) = 16.75%，2020年非活跃人群共有7706602人，活跃人群共有2213476人，活跃人群占总人数的比例为2213476 / (7706602+2213476) = 22.31%。从聚类结果可以看出，从18年到20年，活跃人群的比例从16.75%增加到了22.31%。

综上所述我们可以得到二维码刷码过闸这项地铁新技术的提出对居民出行选择产生了积极影响，该技术的使用促使了更多的居民选择乘坐地铁绿色出行。

1. 居民出行方式选择行为模型构建
   1. 居民出行调查
      1. 思路设计

为了研究居民对是否会乘坐地铁的选择意向，本文需要实证研究的方法。具体的研究思路如下：首先，基于之前有关公交车与居民出行选择关系的研究成果以及居民出行选择因素的研究成果之上来设计本文的调查问卷。通过发放问卷进行居民满意度调研，并利用SPSS软件将所得数据处理成能够建模的形式，最后再对模型进行探索性分析和验证性分析，得出相应结论。

通过前人的研究成果能够总结出与地铁新技术相关的4条性质，以及居民出行选择的内生变量和外生变量，归纳总结，为本文问卷的设计提供参考。

通过了解构建模型和设置指标的理论基础来明确需要得到哪些自变量和潜变量，从而确定问卷设计结构的合理性和科学性。

而问卷则分为预调查以及规模研究二种方法,预调查是为了更好的完善调查结果,更为清楚本文中所要研究的信息以及让被调查者能更为了解问题以便给出更符合标准的回答,而本文所调查问卷信息的主要来源则是规模研究。而最后,本文将会基于SPSS20.0完成所有的大数据分析和模拟研究。具体的数据分析过程主要包括:通过SPSS,首先对所获取的数据采用二元Logistic回归方法,来研究城市四项中轨道交通新技术的特性和市民是否会使用轨道交通出行这一判断间的关联,接着再对内生变量和外生变量采用控制变量分解以说明变量间的相互作用,最后再通过结构方程模型对数据加以验证性分析, 进而得到结论。

* + 1. 问卷设计

问卷设计的目的是通过对居民对地铁满意度的调查来建立模型研究本文的研究假设，得出地铁新技术与居民出行选择相关的结论。问卷设计的质量高低直接决定了后面建模的难易程度以及所得结论的有效性。

本次的问卷主要包括三个部分，第一部分主要是出行者的个人信息，包括性别，年龄，职业。第二部分则是除了地铁新技术之外影响出行者选择的外部因素，包括天气、景区、居民住宅区和是否为交通枢纽，最后部分则是居民对于地铁新技术四个性质的满意程度，包括有形性、可靠性、响应性和保证性，并且将四个方面进行了具体化描述有助于被调研者理解并做出选择。

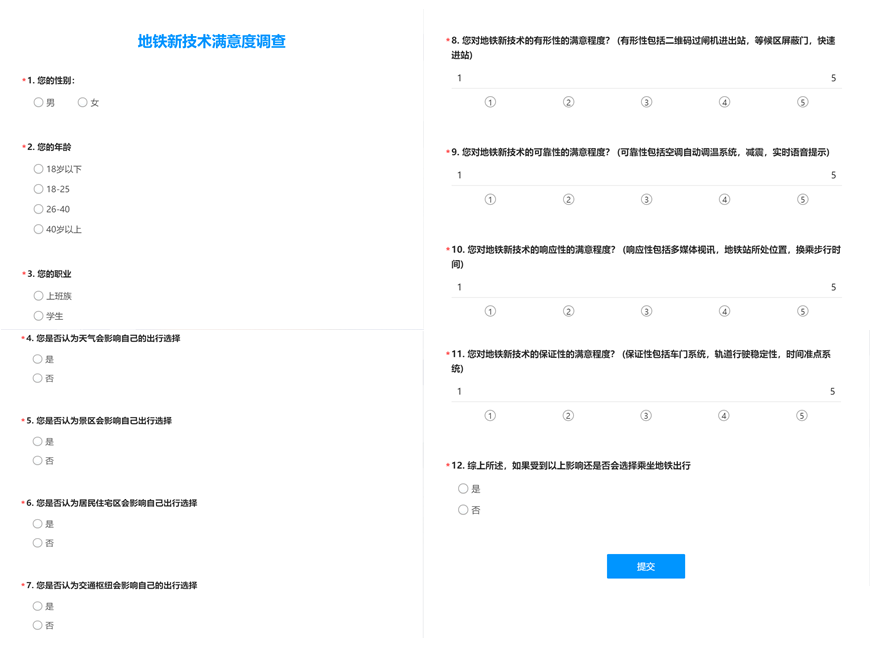


图4-1 调查问卷展示

* + 1. 居民出行调查问卷数据整理

居民出行调查问卷数据如下图。

图4-2 调查问卷数据

具体变量解释如下表。

表4-1 外生变量的具体解释

|  |  |  |
| --- | --- | --- |
| 类别 | 变量名称 | 变量符号 |
| 外生变量 | 性别 | 男=1，女=0 |
| 年龄 | “18岁以下”=1，“18-25岁”=2，“26-40岁”=3，“40岁以上”=4 |
| 职业 | “上班族”=1，“学生”=0 |

表4-2 内生变量的具体解释

|  |  |  |
| --- | --- | --- |
| 类别 | 变量名称 | 变量符号 |
| 内生变量 | 天气是否影响 | 是=1，否=0 |
| 景区是否影响 | 是=1，否=0 |
| 居民住宅区是否影响 | 是=1，否=0 |
| 交通枢纽是否影响 | 是=1，否=0 |

表4-3 模型变量地铁新技术性质的具体解释

|  |  |  |
| --- | --- | --- |
| 类别 | 变量名称 | 变量符号 |
| 地铁新技术性质 | 有形性满意度 | “非常不满意”=1，“比较不满意”=2，“中立”=3，“比较满意”=4，“非常满意”=5 |
| 可靠性满意度 | “非常不满意”=1，“比较不满意”=2，“中立”=3，“比较满意”=4，“非常满意”=5 |
| 响应性满意度 | “非常不满意”=1，“比较不满意”=2，“中立”=3，“比较满意”=4，“非常满意”=5 |
| 保证性满意度 | “非常不满意”=1，“比较不满意”=2，“中立”=3，“比较满意”=4，“非常满意”=5 |

由于疫情原因，本次调查主要以线上发放链接和分享二维码的形式发放问卷，通过历时3周的问卷收集，最终共得到了402份问卷，通过对问卷数据进行完整性和有效性的筛选，最后实际回收的有效问卷有356份，占总获得问卷的90.8%。在此次的调查中，就性别而言，男性有154人，占样本的43.3%，女性有202人，占样本的56.7%。就年龄而言，18岁以下的只有2位，占0.6%，18-25岁的有248份，占样本总数的69.7%，26-40岁的有42份，占11.8%，大于40岁的有64人，占18.0%。就职业而言，学生有298人，占样本总数的83.7%，而上班族的有58人，占样本总数的16.3%。并且几乎所有人都认为天气、是否为景区、是否为居民住宅区、是否为交通枢纽会影响自己出行选择。而关于地铁新技术的有形性、可靠性、响应性和保证性方面，大部分人的评分为3分及3分以上，即代表中立或者满意的态度，由于不同城市的地铁新技术发展程度和政策落实存在差异性，因此少部分被调研者对于自己不了解的地铁新技术评分较低。

* 1. Logistic回归模型分析
     1. 模型的选择

首先，先对模型进行引入。本文中假设 “居民是否选择乘坐地铁出行” 为一种, 连续反应变量代表了某事件出现的可能性, 值域为全体实值, 当某变量的值范围超过了一个临界值m时, 该事件就会出现, 该事件将会发生，即 “居民会选择乘坐地铁出行”，有下列式子[47]：

(4-1)

在这里面， 通常是指实际观测到的反应变量。表示事件将会发生， 表示事件不会发生。如果假设该反应变量和自变量之间存在线性关系，即[47]

(4-2)

则能得到[47]

P(=1|) = P[()>0] = P[()] (4-3)

通常，假设上述式子中有Logistic分布。因此可以将式子改写成[47]：

P(=1|) = F( P[] = (4-4)

该函数就是最基本的Logistic回归函数，成S型分布[47]。

因为在解析变量时，Logistic返回模式一般使用对数线性模型，所以Logistic模式又叫做逻辑回归模型[47]。Logistic是一个广义的探索性线性回归模式，描述的是y取每个值的概率p与x的关系，所以y的每一个取值都会有一个相对应的Logistic回归模型。又因为概率之和等于1，所以k值变量的因变量只需要训练k-1个回归模型[47]。Logistic模型的因变量为定性变量，根据因变量y可能取到的结果，可分为二项逻辑回归和多元逻辑回归[47]。Logistic回归分析方法需要因变数的类型可变，包含顺序变量和名义变量。不管哪种数据类型的变量，都要转换为具体的数值形式来说明它的取值。

在本研究中，Logistic回归通用模型如下式4-5[47]：

p=p(y=1|x) = (4-5)

对Logistic回归模型的概率p进行logit变换如下式4-6[47]：

Logit(p) = ln (4-6)

得到Logistic 回归模型的另一种形式，它给出的是变量 z=logistic ( ) 关于自变量 x 的线性函数[47]：

Logit(p) = ln = (4-7)

推广到多值变量的logistic回归模型，如下式4-8[47]：

Logit(pj) = ln = (4-8)

上述式子也称为发生比率。发生比率可以很明确的解释出自变量对事件发生的机率的影响是正值或者负值。当按照对照组的时间顺序发生变化时，比率就能够互相变化。不管将变量频数增加至多少倍，发生比率都不会产生变化。因为发生比例既可以作为更多分类变量，又可作为更多元模型[47]。

由于本章要探讨的市民出行数据中，最后即市民是否会决定搭乘地铁所取得的数据有选择搭乘与不选择搭乘二类，所以本章使用的是二项逻辑回归模式。二项Logistic回归模拟，是一种研究分析自变量与影响结果的因子间关系的统计分析方式，通过把若干个单独的自变量与因变数或者对单个因变量进行关联研究，分析因变量从而对因变量做出较为准确的预测[47]。这种方法需要对每个自变量都分别引入模型，实现Wald验证后，再继续回归，

Wald = (4-9)

其中，S.E.是指回归系数标准差，B是指第n个自变量的回归系数。假设和已知的显著程度相比较，Wald的统计值的可能性越小，就越说明自变量和该因变量之间存在着显著关系，此自变量应该保留。如果该自变量与因变量无显著关系，那该自变量可以不被考虑[47]。

二项Logistic回归分析一般要求纳入的样本数量为自变量个数的20倍以上，本文所研究的自变量影响因素一共12个，样本数量为356，因此能够使用二元Logistic回归来对居民是否会选择乘坐地铁出行的影响因素进行分析。具体过程为：运用SPSS统计分析软件中的分析 → 回归 → 二项 Logistic模块进行二项Logistic回归分析，如图所示[47]。

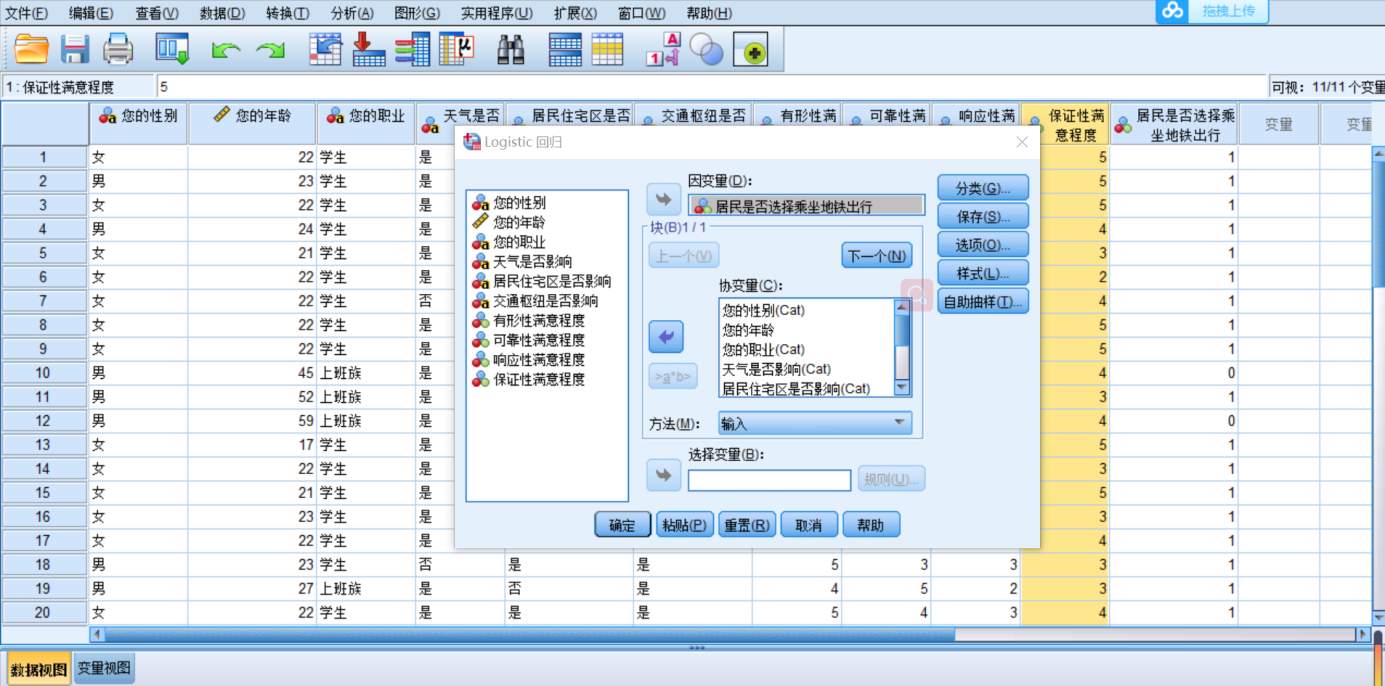


图4-3 建立Logistic回归

回归系数标准误差是指对误差项标准差的估计值，平方为误差项的无偏估算平均值。在回归方程中，表示不同自变量x对因变数y的影响程度的指标参数[47]。回归系数越大，表示x对y的影响越大，正回归系数表示y随x的增大而增大，负回归系数表示y随x增大而减小。z值是指其中一个特征值与均值之间的标准偏差数量，能够观测统计量与假设总体参数之间的差值，可以用z值来确定是否要否定原假设。Wald统计值是用来确定某个变数是不是已经包括到该模式中，或者近似于自由度等于参变量个数的卡方分布， 同时也能够透过比较Wald统计值的多少，来考虑某一自变数在回归模式汇总中的重视程度。Wald的统计值越大，显著性就越高，证明该自变量越重要。p值是通过HosmeR和Lemesshow的拟合优度检验区，通过对观察频数和预期频数之间的卡方检验的量，进而通过模式的设定指标对Logistic的检测。一旦p值等于了给定的明确性水平量，则一个因变数的观察值和模型预测值有差别的零假设，从而说明模式的估计值存在显著性。而一旦p值超过了给定的明确性水平量，则可以接受模型的估计并能够拟合数据。而Cox和Snell的R2则是通过一种在似然标准差的基本上，模拟线性返回的R2求解Logistic回归模式，值通常低于1。NagelkeRke的R2是把Cox和Snell的R2进行了进一步改变，使Cox和Snell的R2加上了自己的最大值。

* + 1. 模型的建立

指定分类数据变量的Gender、Age、Work、Live、Income、License的虚拟变数产生规律为Indicator，数据变量的筛选策略确定Forward：Wald，在使用SPSS软件对调查数据进行二元Logistic回归，之后可以得出表4-4和表4-5中所示结论。

表4-4 二元Logistic回归分析A

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 项 | 回归系数 | 标准误差 | z值 | Wald x2 | p值 | OR值 | OR值95% CI |
| 性别 | 0.094 | 0.327 | 3.754 | 10.683 | 0.000 | 0.910 | 0.480~1.727 |
| 年龄 | 0.104 | 0.254 | 0.154 | 0.167 | 0.683 | 1.109 | 0.675~1.824 |
| 职业 | 0.580 | 0.501 | 3.414 | 10.341 | 0.000 | 0.560 | 0.210~1.494 |
| 有形性 | 0.202 | 0.650 | 5.412 | 16.125 | 0.000 | 0.817 | 0.000~null |
| 可靠性 | 0.507 | 0.512 | 4.121 | 12.456 | 0.000 | 0.602 | 0.000~null |
| 响应性 | 0.052 | 0.350 | 4.367 | 13.562 | 0.000 | 0.950 | 0.000~null |
| 保证性 | 0.253 | 0.472 | 2.561 | 7.541 | 0.021 | 1.288 | 0.000~null |
| 截距 | 4.496 | 1.425 | 3.154 | 9.950 | 0.002 | 89.679 | 5.487~1465.607 |

因变量：居民是否会选择乘坐地铁出行

表4-5 二元Logistic回归分析B

|  |  |  |
| --- | --- | --- |
| McFadden R2 | Cox & Snell R2 | Nagelkerke R2 |
| 0.229 | 0.224 | 0.242 |

从上表可知，将性别，年龄，有无职业，有形性，可靠性，响应性，保证性共7项为自变量，而将居民是否会选择地铁出行作为因变量进行二元Logistic回归分析。

模型公式为：ln = 4.496 + 0.094\*性别 + 0.104\*年龄 + 0.580\*有无职业 + 0.202\*有形性 + 0.507\*可靠性 + 0.052\*响应性 + 0.253\*保证性 (其中p代表居民是否会

选择地铁出行为1 的概率，1-p代表居民是否会选择地铁出行为0的概率)。最终具体分析可知：

性别的回归系数值为0.094，但是呈现出显著性(z=3.754，p=0.000<0.1)，意味着性别会对居民是否会选择地铁出行产生影响关系。

年龄的回归系数值为0.104，但是呈现出显著性(z=0.154，p=0.683<0.1)，意味着年龄会对居民是否会选择地铁出行产生影响关系。

有无职业的回归系数值为0.580，但是呈现出显著性(z=3.414，p=0.000<0.1)，意味着有无职业会对居民是否会选择地铁出行产生影响关系。

有形性的回归系数值为0.202，但是呈现出显著性(z=5.412，p=0.000<0.1)，意味着有形性会对居民是否会选择地铁出行产生影响关系。

可靠性的回归系数值为0.507，但是呈现出显著性(z=4.121，p=0.000<0.1)，意味着可靠性会对居民是否会选择地铁出行产生影响关系。

响应性的回归系数值为0.052，但是呈现出显著性(z=4.367，p=0.000<0.1)，意味着响应性会对居民是否会选择地铁出行产生影响关系。

保证性的回归系数值为0.253，但是呈现出显著性(z=2.561，p=0.000<0.1)，意味着保证性会对居民是否会选择地铁出行产生影响关系。

总结分析可知：性别，有无职业，有形性，可靠性，响应性，保证性这6项都会对居民是否会选择地铁出行产生影响关系。根据z值大小从高至低排序，前三名分别为有形性、可靠性和响应性。因此这三个自变量的显著性最高，对居民是否选择乘坐地铁出行的影响最大。根据Wald x2值的大小从高至低排序，前三名分别是有形性，可靠性和响应性。Wald x2值越大，显著性越高，证明自变量越重要。因此有形性，可靠性和响应性这三项自变量最重要，对居民是否会选择乘坐地铁出行这一潜变量的影响最大。

* 1. 结构方程模型分析
     1. 模型的选择

构造方程模型是通过变和量协方差矩阵来解析变数间关联关系的一类统计分析方法是，多元统计分析中的重要工具。构造方程模型也是一个构建、预测和检验因果关系模式的重要具体方法。模式中包含了可观察的显性变数和隐含的潜变量。结构方程模型能够代替多重回归、通径分析、因子分析和协方差等传统理论数据分析，来解析单个指标对总体结构的影响以及与单个技术指标内部的互相关联。而结构方程模型又可分成计量建模和构造建模。计量建模是指数与潜变量相互之间的关联，而构造建模则是指与潜变量间的关联。

结构方程组解析中能够同样考虑和解决多种因变量。在回归分析中，将会对各种因变量逐一统计，并在结果表格中显示各个因变量。并且态度和行为等变量存在误差。而由于结构方程分析中容许的自变量与因变量之间均含计算误差，因此变量数可用多种指标衡量。在本文中，可以使用因子分析对各个潜变量统计潜变量和二维码支付之间的相关性，得出作为潜变量观测值最大的因子，之后再统计因子分数作为最小的相关系数值。在结构方程研究中，也可以用各个模型对多个样本数值的整体拟合程度来表达问题[50]。

结构变量间的内部结构关联可使用一连串的线性方程来解决，叫做内部结构微分方程。一般是定义了所假定的外生结构变量与内生结构变量间的关系，一般写成以下式4-10[50]：

(4-10)

其中[50]，

η=（η1，…, ηp）T—内部结构变量，其值在方程内部确定的结构变量；

=（ξ1,…,ξq）T—外生结构变量，其值在方程之外确定的结构变量；

=（βij）p\*p—内生结构变量之间的关系；

Г=（γij）p\*q—外生结构变量对内生结构变量的影响；

ζ=（ζ1，…，ζp）T—结构方程的残差项，反映了η在方程中未能被解释的部分。

在本文中，潜变量只有一个：居民是否选择地铁出行；外生变量从天气，景区，居民住宅区以及交通枢纽四个方面进行讨论；内生变量有4个，分别为可靠性η1、有形性η2、响应性η3、保证性η4。可得到如下结构方程[50]：

**=** (4-11)

本文在研究居民是否选择乘坐地铁时引入潜变量，利用结构方程模型（SEM）介绍了潜变量与它对应的测量变量间的相应关系。结构方程模型能够分析出单个或多个不同的自变量与单个或多个不同因变量间存在的相互关系，同时对于自变量和因变量的限定较少，能够连续也能够间断。

每个设计变量和观测变量间的外部相互作用可以使用其他的过程来加以说明，称为测量过程。一般要通过二个方程式来规定内生结构变量η与内生解释变量y之间的关联，以及与外生构造变量x之间的联系，一般写成以下形式[50]：

x = (4-12)

y = (4-13)

其中[50]，

x = (x1… xn)—外生观测变量向量；

y = (y1… ym) —内生观测变量向量；

= (Мij)m\*k—外生观测变量与外生结构变量之间的关系，是外生观测变量在外生结构变量上的因子负荷矩阵；

= (Мij)n\*i—内生观测变量与内生结构变量之间的关系，是内生观测变量在内生结构变量上的因子负荷矩阵；

=（）T—外生观测变量x的误差项；

=（）T—外生观测变量x的误差项；

在进行建立结构方程建模之前，就必须建立路径图来说明结构变量间的直接关联。 一般用矩形框代表可观察变数，用椭圆框代表原结构变数，用垂直箭头来确定结构变量之间的直接关系，箭头从起因变数导向最终结果变数[50]。如果假想结构变量间没有联系，即代表与假想变数之间没有直接关联。而反应性指数和形成性指数则是结构变量和观察变数间，基于因果关系中的优先属性而形成的各种概念的指数关系[50]。若假想的结构变量能够被认为由一个基地所组成，那么通过这种构造而形成的被考察的事物就可以反应这个结构变量的指数，进而形成反该结构变量的指数，从而成为反应性指数。观察结构变量是一个效应，而结构变量则是因子。反之，如果认为结构变量是被观测变量而产生影响则说明，结构变量是效应，而观测变量则是影响因素，且彼此之间是线性关系，而观测变量则是影响形成的指数，如图4-4[50]。

**x2**

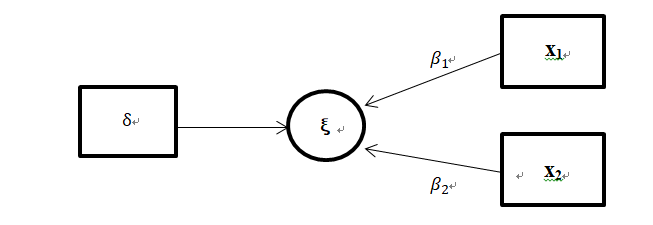
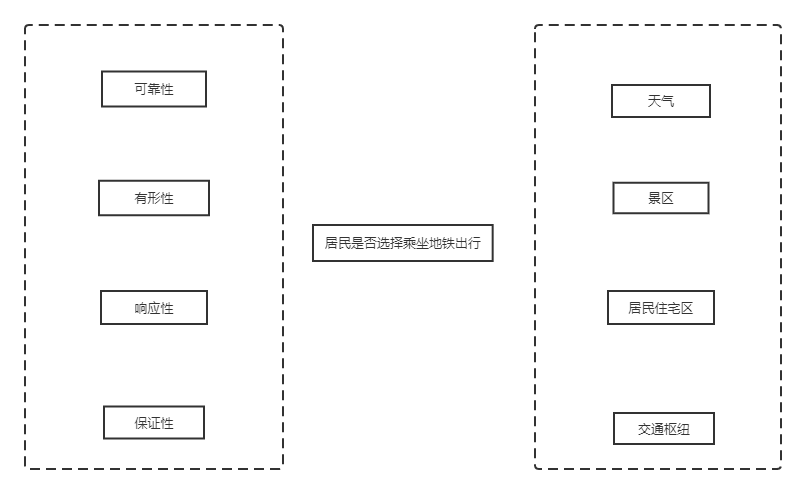


图4-4 形成性指标路径示例

回归方程表示如下[50]： (4-14)

* + 1. 模型的建立

首先是建立模型的形成性指标路径，如图4-5。

图4-5 形成性指标路径

然后，利用Amos软件运行结构方程模型来研究各个因子对居民是否选择乘坐地铁出行的形成性指标路径结果，具体结果如下表4-6。

表4-6 结构方程模型分析

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X |  | Y | 非标准化路径系数 | SE | z（CR值） | p | 标准化路径系数 |
| 交通枢纽 |  | 居民是否会选择地铁出行 | 0.90 | 0.066 | 5.361 | 0.004 | 0.78 |
| 居民住宅区 |  | 居民是否会选择地铁出行 | 0.14 | 0.023 | 3.400 | 0.048 | 0.46 |
| 景区 |  | 居民是否会选择地铁出行 | 0.31 | 0.022 | 3.312 | 0.058 | 0.14 |
| 天气 |  | 居民是否会选择地铁出行 | 0.29 | 0.016 | 1.806 | 0.071 | 0.118 |
| 保证性 |  | 居民是否会选择地铁出行 | 0.75 | 0.053 | 3.375 | 0.051 | 0.78 |
| 响应性 |  | 居民是否会选择地铁出行 | 0.44 | 0.031 | 1.129 | 0.097 | 0.47 |
| 可靠性 |  | 居民是否会选择地铁出行 | 0.19 | 0.041 | 3.672 | 0.037 | 0.26 |
| 有形性 |  | 居民是否会选择地铁出行 | 0.77 | 0.037 | 6.124 | 0.000 | 0.65 |

备注： 表示路径影响关系

从上表可知：交通枢纽对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=5.361，p=0.004<0.1)，因而说明交通枢纽对居民是否会选择地铁出行会产生影响关系。

居民住宅区对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=3.400，p=0.048<0.1)，因而说明居民住宅区对居民是否会选择地铁出行会产生影响关系。

景区对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=3.312，p=0.058<0.1)，因而说明景区对居民是否会选择地铁出行会产生影响关系。

天气对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=1.806，p=0.071<0.1)，因而说明天气对居民是否会选择地铁出行会产生影响关系。

保证性对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=3.375，p=0.051<0.1)，因而说明保证性对居民是否会选择地铁出行会产生影响关系。

响应性对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=1.129，p=0.097<0.1)，因而说明响应性对居民是否会选择地铁出行会产生影响关系。

可靠性对于居民是否会选择地铁出行影响时，此路径呈现出显著性(z=3.672，p=0.037<0.1)，因而说明可靠性对居民是否会选择地铁出行会产生影响关系。

有形性对于居民是否会选择地铁出行影响时，此路有呈现出显著性(z=6.124，p=0.000<0.1)，因而说明有形性对居民是否会选择地铁出行会产生影响关系。

在该模型中显著性最高的三项分别是有形性、可靠性和保证性，该结构方程主要起到验证性作用，进一步确定Logistic回归模型分析得出的结论可靠性。

* 1. 控制变量分析

除了地铁新技术的性质会对居民是否会选择乘坐地铁产生影响外，还有其他因素也会对居民出行选择产生一定的影响，这些因素可能会在一定程度上影响我们的结论。根据结构方程模型理论，我们可以把这些其他因素归纳成内生变量和外生变量两类，并且需要探究这些因素对居民出行选择产生影响的显著程度，从而判断是否可以忽略或者界定。

* + 1. 内生变量与解释变量关系

内生变量包括天气、景区、居民住宅区以及交通枢纽，具体模型结果如图。

表4-7 内生变量分析A

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 项 | 回归系数 | 标准误差 | z值 | Wald x2 | p值 | OR值 | OR值95% CI |
| 天气 | 0.199 | 0.118 | 1.678 | 2.814 | 0.093 | 1.220 | 0.967~1.539 |
| 景区 | 0.231 | 0.173 | 2.332 | 3.776 | 0.043 | 0.794 | 0.566~1.115 |
| 居民住宅区 | 0.083 | 0.192 | 1.730 | 2.985 | 0.087 | 1.086 | 0.745~1.584 |
| 交通枢纽 | -0.657 | 0.587 | -1.821 | 3.116 | 0.062 | 0.518 | 0.164~1.636 |
| 截距 | 4.777 | 2.856 | 1.672 | 2.797 | 0.0941 | 18.799 | 0.440~32082.517 |

因变量：居民是否会选择乘坐地铁出行

表4-8 内生变量分析B

|  |  |  |
| --- | --- | --- |
| McFadden R2 | Cox & Snell R2 | Nagelkerke R2 |
| 0.215 | 0.212 | 0.222 |

从上表可知，将天气，景区，居民住宅区，交通枢纽共4项为自变量，而将居民是否会选择地铁出行作为因变量进行二元Logit回归分析，

模型公式为：ln(*p*/1-*p*)=4.777 + 0.199\*天气+0.231\*景区 + 0.083\*居民住宅区-0.657\*交通枢纽(其中*p*代表居民是否会选择地铁出行为1 的概率，1-*p*代表居民是否会选择地铁出行为0的概率)。最终具体分析可知：

天气的回归系数值为0.199，呈现出显著性(*z*=1.678，*p*=0.093<0.1)，意味着天气会对居民是否会选择地铁出行产生影响关系。

景区的回归系数值为0.231，呈现出显著性(*z*=3.332，*p*=0.043<0.1)，意味着景区会对居民是否会选择地铁出行产生影响关系。

居民住宅区的回归系数值为0.083，但是呈现出显著性(*z*=1.730，*p*=0.087<0.1)，意味着居民住宅区会对居民是否会选择地铁出行产生影响关系。

交通枢纽的回归系数值为-0.657，但是呈现出显著性(*z*=-1.821，*p*=0.062<0.1)，意味着交通枢纽会对居民是否会选择地铁出行产生负影响关系。

总结分析可知：天气, 景区, 居民住宅区, 交通枢纽共4项全部均会对居民是否会选择地铁出行产生影响关系。

* + 1. 外生变量与因变量关系

外生变量包括性别、年龄以及有无职业，具体模型结果如图。

表4-9 外生变量分析A

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 项 | 回归系数 | 标准误差 | z值 | Wald x2 | p值 | OR值 | OR值95% CI |
| 性别 | 0.130 | 0.322 | 1.505 | 3.164 | 0.065 | 0.878 | 0.467~1.649 |
| 年龄 | 0.083 | 0.251 | 0.332 | 0.110 | 0.740 | 1.087 | 0.665~1.776 |
| 有无职业 | 0.639 | 0.497 | 2.521 | 7.151 | 0.009 | 0.582 | 0.199~1.399 |
| 截距 | 2.282 | 0.833 | 2.739 | 7.504 | 0.006 | 9.800 | 1.914~50.170 |

因变量：居民是否会选择乘坐地铁出行

表4-10 外生变量分析B

|  |  |  |
| --- | --- | --- |
| McFadden R2 | Cox & Snell R2 | Nagelkerke R2 |
| 0.208 | 0.207 | 0.212 |

从上表可知，将性别，年龄，有无职业共3项为自变量，而将居民是否会选择地铁出行作为因变量进行二元Logit回归分析，模型公式为：ln(*p*/1-*p*)=2.282+0.130\*性别 + 0.083\*年龄+0.639\*有无职业(其中*p*代表居民是否会选择地铁出行为1 的概率，1-*p*代表居民是否会选择地铁出行为0的概率)。

最终具体分析可知：

性别的回归系数值为0.130，但是呈现出显著性(*z*=1.505，*p*=0.065<0.1)，意味着性别会对居民是否会选择地铁出行产生影响关系。

年龄的回归系数值为0.083，但是呈现出显著性(*z*=0.332，*p*=0.740<0.1)，意味着年龄会对居民是否会选择地铁出行不产生影响关系。

有无职业的回归系数值为-0.639，但是呈现出显著性(*z*=2.521，*p*=0.009<0.1)，意味着有无职业会对居民是否会选择地铁出行产生影响关系。

总结分析可知：性别，有无职业共2项全部均会对居民是否会选择地铁出行产生影响关系。

* 1. 研究结果分析

该部分通过建立Logistic回归模型来对地铁新技术以及其他因素对居民是否选择乘坐地铁出行这一结果进行探究性分析，然后再利用结构方程模型对这两者之间的关系进行验证性分析。根据控制变量可以看出内生因素和外生因素都会影响居民是否乘坐地铁。但是根据模型结果可以看出，地铁新技术与居民出行是否选择乘坐地铁出行有很强的相关性，地铁新技术与居民出行选择的显著程度目前大于其他内生因素和外生因素与居民出行选择的显著程度。因此可以看出比起其他年龄、性别、天气和居民住宅区等其他变量因素，居民更加在乎地铁带来的舒适度、便捷度和安全性等设施条件，这些都与地铁新技术的实施和发展息息相关。该部分主要证实了居民是否乘坐地铁与地铁新技术之间存在了显著关系，更加能够证实第三章中关于地铁新技术政策的实施对居民出行选择有积极影响这一结论。

1. 相关政策建议
   1. 其他地铁新技术政策分析

2003年全国第一套地铁屏蔽门系统在广州地铁二号线正式投入使用。北京市地铁屏蔽门是在2017年才正式全面投入使用。地铁屏蔽门系统能够保护乘客安全、节约环控系统运营能耗、改善站台乘客候车环境的功能，这也为之后的地铁无人驾驶技术奠定了基础。由于数据获取有限，所以其他地铁新技术主要通过观察北京市官网上找到的如下能够显示不同年份的OD图来初步判断是否对居民出行选择有影响，如若得到准确结论还需再做如上述二维码刷码过闸技术例子的具体研究，下面是2017年和2018年地铁屏蔽门使用前后官网OD图如下。

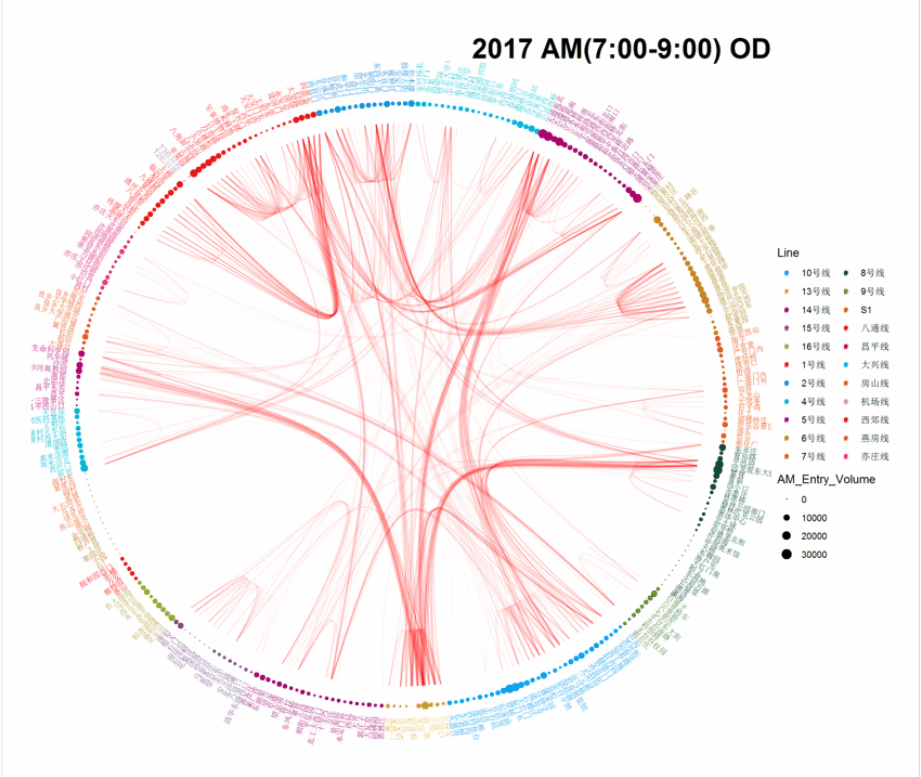


图5-1 2017时间差变化结果图

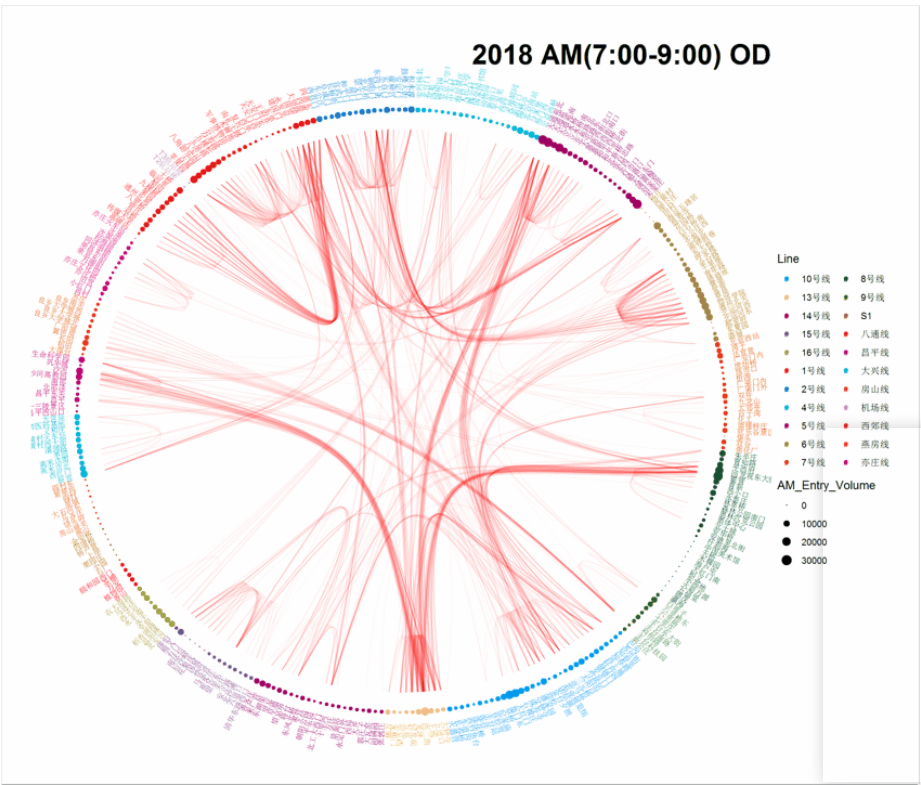


图5-2 2018时间差变化结果图

如图可以观察到2018年的客流量比起2017年的客流量有微小的增长，因此可以后续对地铁屏蔽门这一地铁新技术进行进一步研究来证实本文的普适性。

2016年全国最初的地铁空调通风系统在北京地铁正式投入使用。地铁空调通风系统能够自动适应地铁乘客动态舒适温度，解决了传统热舒适理论对乘客在地铁内不同区域无法准确预测人体热舒适度的难题，对人体在热环境中舒适预测精度提升20%以上。这种设计能够充分利用地铁存在活塞风和热压风的特点，在不影响乘客动态热舒适情况下，用自然风速补偿温度和湿度，从而显著降低空调系统装机容量，达到节约地铁初期建设投资和后期运营维护费用的目的。由于数据获取有限，所以其他地铁新技术主要通过观察北京市官网上找到的如下能够显示不同年份的OD图来初步判断是否对居民出行选择有影响，如若得到准确结论还需再做如上述二维码刷码过闸技术例子的具体研究，下面是2016年和2017年地铁空调通风系统使用前后官网OD图如下。

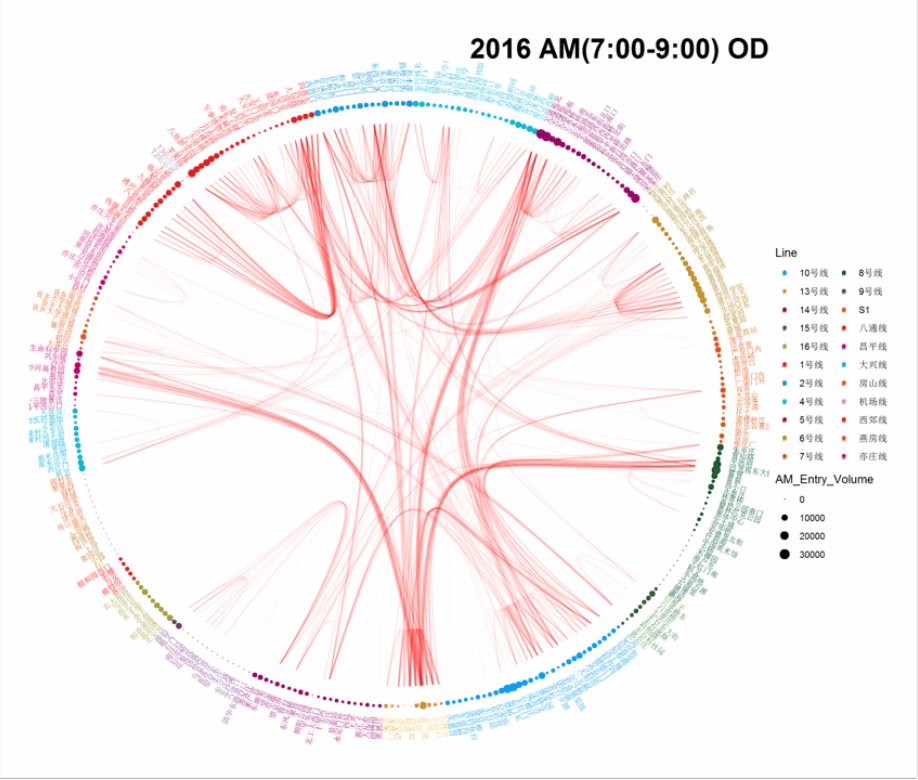


图5-3 2016时间差变化结果图

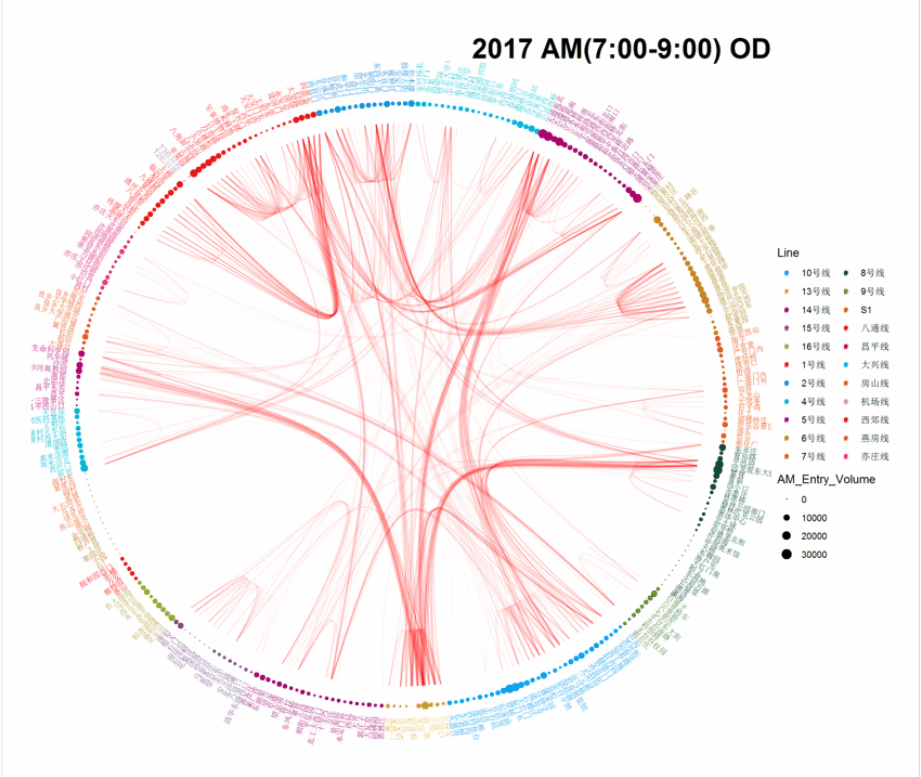


图5-4 2017时间差变化结果图

如图可以观察到2017年的客流量比起2016年的客流量有微小的增长，因此可以后续对地铁空调通风系统这一地铁新技术进行进一步研究来证实本文的普适性。

* 1. 相关政策建议

基于“地铁新技术的使用对居民出行选择产生积极影响”这一结论，本文可以对相关政策提供以下建议：

（1）积极推进并指导“互联网+”在轨道交通行业的应用方向，对科技变革可行性分析提供有力证据。

（2）该结论说明城市居民对地铁新技术具有较高的反应程度和适应程度，因此政府以后可以将政策发布的信息倾向于直接介绍地铁新技术的优点以及使用方法，不再需要过多担心居民的接受程度。

（3）利用该结论还可以响应和实施环保部关于推进绿色出行发展，减少环境污染，建设低碳社会，促进可持续发展战略。

（4）政府可以大力推动与居民出行行为相关的地铁新技术的发展，使该地铁新技术在生活中变得普适化，使居民出行更加便捷。

（5）本研究结论还可以作为环境领域公共政策的发展方向，作为绿色出行领域在轨道交通方面发展的理论基础。

1. 研究结论

首先，由于地铁新技术种类繁多，所以本文的地铁新技术将界定为与居民出行相关的城市地铁系统设备新技术、城市地铁智能化系统新技术、城市地铁规划和设计新技术以及城市地铁运营组织管理新技术这四个领域之中。本文以城市地铁系统设备新技术中的二维码刷码过闸技术为例，通过对二维码刷码过闸政策颁布前后数据的客流量变化、时间差变化和活跃人群变化的研究来探究地铁新技术与居民出行选择之间的关系，得到了以下结论。第一，政策颁布后的客流量比政策颁布前的客流量有微小的增加，但是该结论受到的疫情影响较大，无法得出具体的关系结论。第二，二维码刷码过闸技术使用后的居民进出站时间差小于该技术使用前的时间差，由于该因素的影响可能会促使更多学生、上班族等居民选择乘坐地铁出行。第三，通过对活跃天数和钱包余额聚类分析来衡量活跃人群的活跃天数的聚类变化，钱包余额大的属于活跃人群，钱包余额小的属于非活跃人群。根据对大量数据聚类得出的聚类分析图可以看出，18年活跃人群的活跃天数基本全部小于12天，而20年活跃人群的活跃天数分布在10-17.5之内，可以发现活跃人群的活跃天数明显增加了。再对活跃人群进行聚类计算，可以看出从18年到20年，活跃人群的比例从16.75%增加到了22.31%。因此这些都能够说明该地铁新技术的使用会影响居民出行选择，从而促使更多居民选择乘坐地铁绿色出行。

之后，为了进一步佐证居民是否选择乘坐地铁出行与地铁新技术是否真正存在相关性，本文利用问卷调查的数据进行了进一步的模型建立，通过建立Logistic模型和结构方程模型来探究和验证地铁新技术对于居民是否乘坐地铁出行的是否存在影响，得到了以下结论。第一，通过建立Logistic回归模型和结构方程模型可以得到地铁新技术与居民是否选择乘坐地铁出行有显著的相关性，且地铁新技术的有形性、可靠性和保证性与居民出行选择具有更强的相关性。第二，通过控制变量可以得到其他变量也会对居民出行选择产生一定的影响，但是其显著性小于地铁新技术性质的显著性，也就是说比起其他因素的影响，居民还是更加在乎地铁出行的便捷性、安全性以及舒适性等因素，因此在影响居民是否选择乘坐地铁的影响因素中最为重要的还是地铁新技术因素。这更加证实了前一章证明的地铁新技术的使用会影响居民出行选择这一结论是成立的。地铁新技术（比如二维码刷码过闸）的使用确实能对居民出行方式的选择产生积极影响，从而促使居民选择更加绿色的公共交通工具（比如地铁）作为出行方式。

参考文献

1. 张鹏,王健,吴娟,张宁.南京地铁移动支付关键技术的研究及应用[J].都市快轨交通.2020.33(06).146-150+161.
2. 许巧祥,李灏,徐建国,李延伟.南京地铁移动支付的模式选择与战略思考[J].都市快轨交通.2019.32(02).103-108.
3. Geng, J.C., Long R.Y., Chen, H., Li, W.B. Exploring the motivation-behavior gap in urban residents’ green travel behavior: A theoretical and empirical study [J]. Resources.Conservation and Recycling.2017.125.282-292.
4. Delis M D, Iosifidi M. Environmental awareness, consumption, and labor supply: Empirical evidence from household survey data [J]. Ecological Economics.2016.129.1-11.
5. Yuan Y, Zhang Q, Mao A, et al. A study on the Current Situation and Countermeasures of Green Travel for College Students[J]. IOP Conference Series Earth and Environmental Science.2020.531.012053.
6. Zhang W, Tian Z, Zhang G, et al. Spatial-temporal characteristics of green travel behavior based on vector perspective[J]. Journal of Cleaner Production.2019. 234.
7. Zheng J, Xu M, Li R, et al. Research on Group Choice Behavior in Green Travel Based on Planned Behavior Theory and Complex Network[J]. Sustainability. 2019.11.
8. 朱万山,刘兴刚,崔永玲,李鹏,李瑞艺,杜金鹏,刘铁根.城市公共交通中非接触支付技术研究进展[J].电子测量技术.2020.43(09).121-126.
9. 辛欣.大数据和人工智能发展中的智慧地铁运维研究[J].科学技术创新.2020(11).58-59.
10. 齐航.地铁BAS系统的新技术[J].科技创新导报.2015.12(22).106-108.
11. 卢光霖.广州地铁的新技术应用与创新[J].都市快轨交通.2004(01).6-8+20.
12. 柳颖,周慧娟.城市轨道交通乘客出行行为特征分类研究[J].交通工程.2019.19(03).72-77.
13. 傅鹏明.城市居民出行方式选择影响因素分析[C].协同发展与交通实践——2015年中国城市交通规划年会暨第28次学术研讨会论文集.2015.219.
14. Yang M. Multimodal Travel Mode Imputation based on Passively Collected Mobile Device Location Data. 2020.
15. 刘杰,常兴山,孙锋,周建辉.基于聚类分析的数据文件格式分析方法[J].武汉理工大学学报.2022.44(01).93-99.
16. 周畅.K-means聚类分析技术在教育评价中的应用[J].集成电路应用,2022,39(01):126-127.
17. M.Nauta, D.Bucur, C.Seifert. Causal discovery with attention-based convolutional neural net-works[J]. Machine Learning and Knowledge Extraction.2019.1(1).312-340.
18. 李祥.轨道交通乘客出行模式及客流演化[D].北京交通大学.2021.
19. Luan X, Cheng L, Wu Z. Residents’ action response experiments on different travel modes under the external conditions of a metropolitan environment[C].IOP Conference Series:Earth and Environmental Science. IOP Publishing Ltd.2020:032057 (7pp).
20. 耿纪超,龙如银,陈红.居民出行方式选择影响因素的研究述评[J].北京理工大学学报(社会科学版).2016.18(05).1-9.
21. 何瑞春,李引珍,张峻屹,藤原章正.城市居民出行选择预测模型及实证研究[J].交通运输系统工程与信息.2007(06).80-84.
22. Geng, J.C., Long R.Y., Chen, H., Yue, T., Li, W.B., Li, Q.W. Exploring Multiple Motivations on Urban Residents’ Travel Mode Choices: An Empirical Study from Jiangsu Province in China [J]. Sustainability.2017.9(1).136.
23. Srinivasan S, Rogers P. Travel behavior of low-income residents: studying two contrasting locations in the city of Chennai, India[J]. Journal of Transport Geography.2005.13(3).265-274.
24. Lin B, Wang X. Does low-carbon travel intention really lead to actual low-carbon travel? Evidence from urban residents in China. 2021.
25. Ao Y, Zhang Y, Wang Y, et al. Influences of rural built environment on travel mode choice of rural residents: The case of rural Sichuan[J]. Journal of Transport Geography.2020.85.
26. 耿纪超.多元动机视角下城市居民出行方式选择及其引导政策研究[D].中国矿业大学.2017.
27. Rya B, Lla B, Jza B. Impact of the consciousness factor on the green travel behavior of urban residents: An analysis based on interaction and regulating effects in Chinese cultural context[J]. Journal of Cleaner Production.2020.274.
28. Liang L, Xu M, Grant-Muller S, et al. Household travel mode choice estimation with large-scale data—an empirical analysis based on mobility data in Milan[J].International Journal of Sustainable Transportation.2019(4).1-16.
29. Xu Z, Shao C, Wang S , et al. Analysis and Prediction Model of Resident Travel Satisfaction[J]. Sustainability.2020.12(18).7522.
30. 彭志林.二维码车票支付在轨道交通领域的应用及风险对策[J].电子技术与软件工程.2021(08).11-12.
31. 边驰.城市公共交通支付方式影响研究[D].浙江大学.2020.
32. 杨义锋.基于城市公共交通二维码支付系统的研究与实现[D].广东工业大学.2018.
33. 邓进.城市轨道交通车站客流特征与周边土地利用互动关系[D].北京交通大学.2015.
34. Kwon, Soon-Bark, Namgung, et al. Size distribution analysis of airborne wear particles released by subway brake system[J]. Wear: an International Journal on the Science and Technology of Friction.Lubrication and Wear.2017.372/373.169-176.
35. Rohmawati T, Kartiwan I. The Influence of Information Technology on Public Transportation[J]. IOP Conference Series Materials Science and Engineering.2020.879.012004.
36. 宋燕.基于Logistic回归和Probit回归的个人信用评估研究[D].广西师范大学.2019.
37. 尹静.北京市居民出行方式选择意向研究[D].北京交通大学.2012.
38. O.Goudet, D.Kalainathan, P.Caillou, et al. Learning functional causal models with generative neural networks[M]. Springer.2018.39-80.
39. Li C, Shimizu S.Combining Linear Non-Gaussian Acyclic Model with Logistic Regression Model for Estimating Causal Structure from Mixed Continuous and Discrete Data[J]. 2018.
40. 刘昕.基于结构方程模型的因果分析算法研究及应用[D].电子科技大学,2019.
41. 梁嘉.基于结构方程模型HEYDAY品牌延伸影响因素研究与应用[D].北京交通大学,2021.
42. Shah B A, Zala L B, Desai N A. Structural equation modelling for segmentation analysis of latent variables responsible for environment-friendly feeder mode choice[J]. 2022.
43. Li W, Zhao S, Ma J, et al. Investigating Regional and Generational Heterogeneity in Low-Carbon Travel Behavior Intention Based on a PLS-SEM Approach[J]. Sustainability.2021.13.
44. 许熳灵.基于地铁刷卡数据的城市居民出行行为时空特征分析[D].东南大学.2019.
45. 赵娟娟.城市轨道交通乘客时空出行模式挖掘及动态客流分析[D].中国科学院大学(中国科学院深圳先进技术研究院).2017.
46. Zhang H, Zhuge C, Jia J, et al. Green travel mobility of dockless bike-sharing based on trip data in big cities: A spatial network analysis[J]. Journal of Cleaner Production.2021.313(1).127930.
47. 宁静,王云,马超群.基于Logistic模型的地铁乘客进站方式选择影响因素分析[C].2019世界交通运输大会论文集（上）.2019.876-884.
48. 马晖玮.基于非集计模型的居民出行方式选择影响因素研究[D].南京财经大学.2015.
49. 周兆丁,吕锟,沈瑾,邹文生,苏子钦,邱巧勇,李明华.统计软件SPSS相关分析及应用[J].电脑知识与技术.2019.15(20).301-302.
50. 冯秀琴.基于结构方程模型的顾客满意度测评与优化研究[D].武汉理工大学.2009.
51. 王蔚.基于结构方程模型的政策效应评估方法研究[D].北京交通大学.2021.

致 谢

6月是一座无形于天际的洛溪山，山这边是北交的海，山那边是宇宙的茫。5点钟是一朵盛开在人间的火烧云，昨天滑过山顶，明天飘向远方。她以后才会明白，有些再见，便是再也不见。在这场美丽的修行里，一草一木都有灵，一词一句皆有情。

我听见回声，来自山谷与云间。来时，是一潭挣扎的枯藻，带着凄微的北风，穿过失血的心脏；去时，头置簪花，带着驻守岁月的信念，一路盛开。她说，以四年的镰刀收割交织的灵魂，疼痛后便是往复的幸福，就像沙漠里摇曳的绿洲，不凋不败，充盈着纯净和热烈，不盛不乱，贯穿着清骨和傲然。她不悲不喜，无忧无伤，也仍是无畏离合变迁，流年兮转，只是如今也会时不时的停留在老窗前，听静水回风，听花败花开。她忘记自己的年龄，忘记过往的遗憾，忘记困扰的心结，忘记所有的悲，只于心里留下善因，是他，是她，是她们，是童年，是青春，是陪伴，是帮助，是未来，是从此以后都希望能报之以歌，温柔生活。

燕子南飞，浪迹天涯的小舟，遗世独立的梅花，渡口日落烟霞，都留在了古早的时间里，林下相逢，从不问因果，只会永远留存于心间，如四月的春风。

附 录

附录A Python运行过程图

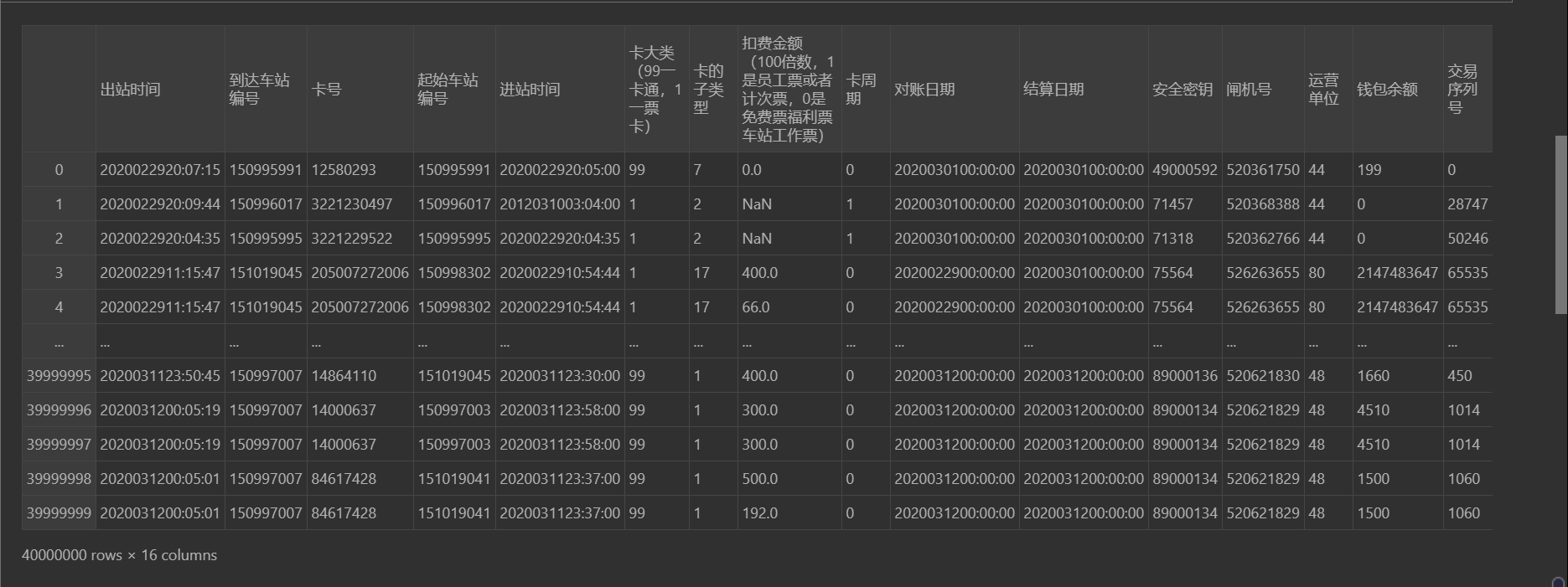


图1 2020年整合数据读取过程

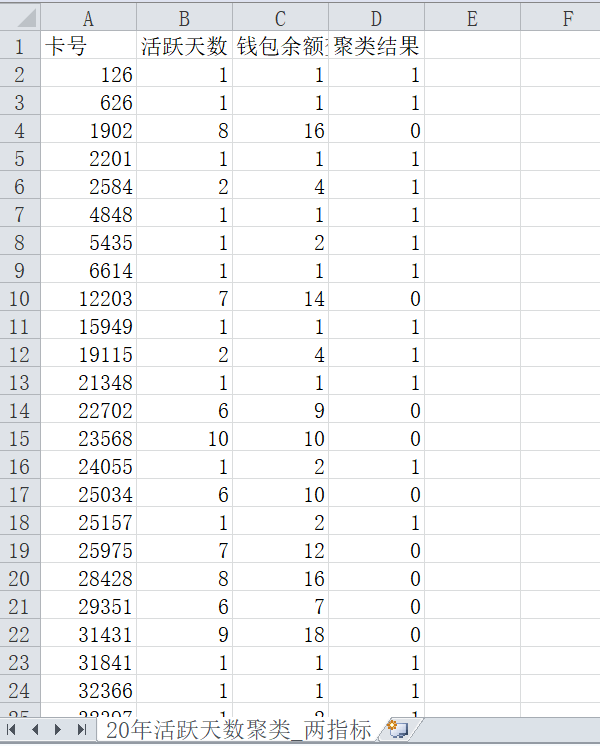


图2 2020年时间差数据

图2 2020年时间差数据

