公交移动支付问题的评估方案

摘 要

本文参考附件数据并查阅大量资料、数据，构建了相应的数学模型，对乘客出行行为特征、平台盈利模型和未来的商业可行性等问题进行了研究。

对于问题一，首先对数据进行预处理：读入数据，对数据进行整合后，筛除未刷卡以及刷卡未记录的异常数据。然后，将出行支付基础特征如乘车时间、支付方式、乘车人作为基本因素利用 Pandas 进行数据分析、数据可视化。在分析了 24h、一周内以及月份间的支付方式，并且将乘客分类比较同类乘客的倾向性选择，以及不同乘客在不同时间的支付方式选择概率后，可得知该城市乘车人在不同时段内的支付方式：公交卡支付次数略大于移动支付，但均接近 50%；同类乘客在面临多种支付方式时，选择某种方式的概率也大致相同；乘车人在选择支付方式时基本不会长时间倾向于某种支付方式。

对于问题二，首先进行第三方支付平台盈利模式的探究：通过查阅相关资料结合公共交通的特性，分析得出平台具有的 4 种收入途径：广告费、服务费、手续费和沉淀资金利息收入，以及 3 种支出途径：前期广告宣传费用、移动端接入费用、固定支出。在建模过程中，由于变量对应的比例系数并不相同，需要利用分段函数进行线性规划。另外，用户数量、平台影响力与金额之间仍存在隐函数关系，需进一步讨论。通过已有数据和资料，进行未知数据的预测，将其代入盈利模型，定量计算并分析盈利的状况：

$$W_0 = I_0 - O_0 = \left(810 + 5.28 + 100 + 0 - 600 - 39.6 - 85\right) \times 10^4 = 1906800(yuan / month)$$

对于问题三，首先借助问题一的结论，在假设“安装了移动支付设备的四分之一公交、地铁为该城市最热门的线路”的前提下，依据公交线路客流量数据得到已安装设备车辆中移动支付所占比率。然后利用“供求关系”来分析提高设备覆盖率后的变动趋势，进而使用集计（创新扩散模型）与非集计（离散选择模型）两种模型，预测全覆盖条件下，移动支付与公交卡支付所占比率分别为 86.75%与 13.25%，最后利用问题二模型计算得到第三方平台盈利状况，结果为每月盈利约 724.179 万元。

对于问题四，首先查阅资料研究第三方移动支付平台的商业规划与发展状况。结合本次问题中已有的盈利模型进行具体分析。针对与公共交通系统进行合作的模式，研究其盈利状况和平台影响力的增长。通过具体的数据，得出相应的商业规划及发展的可行性方案，从中进行分析进而得出行为特征，并归纳成可行性方案的建议。

总之，本文主要利用 Python 的 Pandas 数据分析模块、MATLAB 等进行编程，归纳了出行支付特征，建立求解了第三方平台收支盈利模型，预测移动支付设备全覆盖条件下第三方平台盈利状况，并进行可行性研究，较好的解决了问题。

关键词： 数据处理与分析 线性规划 创新扩散模型 离散选择模型

目录

一、问题重述

随着智能手机的普及和移动支付技术的提高，现有的现金缴费和实体公交卡刷卡的付费方式存在缺点，如公交卡在使用过程中存在着充值不方便、容易丢失、刷卡记录个人无法查看、跨地区无法使用等问题，现金支付带来很多不便，增加人工成本等，而公交移动支付以其相比现金支付和实体公交卡刷卡的优势逐渐进入我们的视线。

目前公交移动支付一般是通过第三方支付平台进行，现需要你对该项目进行分析和评估。

问题 1：附件 1,2 中给出了某城市的部分公交支付的信息和数据说明，试分析该城市乘车人的出行支付特征。

问题 2：建立一个公交第三方支付平台的商业盈利数学模型，定量分析公交第三方支付平台的收支和盈利情况。

问题 3：问题 1 中给出的数据为四分之一的公交车和地铁安装移动支付设备后试营运期间得到的数据，根据问题 1 中的数据，试估计该城市全部公交实现公交第三平台致富后的盈利情况。

问题 4：结合前面的计算结果和结论，给移动支付公司写一份 500 字以内的商业计划可行性报告，并给出增加公司盈利的可行性方案建议。

二、问题假设

1. 假设安装了移动支付设备的四分之一公交、地铁为该城市最热门的线路。
2. 不考虑除移动设备支付、公交卡支付外的其他支付方式。
3. 假设用户选择是理性的。

三、符号说明

序号	符号	意义
1	W_0	模型的总盈利
2	I_0	模型的总收入
3	O_0	模型的总支出
4	r	沉淀资金利息增长系数
5	φ	用户人数的预测系数
6	σ	阈值的修正误差值
7	X_c	宣传费用的算术平均值
8	N	用户人数的阈值

四、问题分析

4.1 问题一的分析

 为了分析该城市乘车人的出行特征，我们从七个方面对数据进行了不同层面的分析。首先根据数据给出的信息，将乘车时间，乘车人 ID，支付方式等信息分类组合并进行数据处理。分析一天内、一周内、不同月份间的支付次数累和来表明不同时间下人们会选择何种支付方式，将出行的人群分类，并且分析不同人群对于支付方式的倾向性。在相同的条件下，筛选出不在我们讨论范围的异常数据，利用 PANDAS 读入数据，将结果可视化得到数据表格，再进行对数据的分析，分析可能出现的原因以及出行特征。

4.2 问题二的分析

为了得到第三方盈利模型，先要明确第三方支付平台需要在收入和支出的项目，收入项目分为：广告费、服务费、手续费和沉淀资金利息收入。支出项目分为：前期广告宣传费用，移动端接入费用，固定支出（基础设施投入，新项目固定成本，员工工资等）。因为该盈利模型涉及到很多变量，又存在着不随着变量改变的一些其他量，因此将等式列出来后，我们利用行列式，幂函数等函数，将变量细化，最后查找资料设定一些定量的取值，得到最终定量分析的结果，算出公交第三方支付平台的盈利模型能够很好地吻合题目。

4.3 问题三的分析

由于题目中给到的只有四分之一的公交和地铁安装移动支付设备后得到的数据，为了能够建立一个能够推出全部情况的模型，我们需要将占有率由四分之一推至全部，由于第一问的结果，提出这四分之一的线路占有绝大多数客流的假设，并且论证假设成立。将这个结果带入供应-需求模型，用以计算客户需求与第三方支付完全推广之后的关系。然后用著名的经济学家罗杰斯提出的创新扩散模型，将关系推广扩散，推测当第三方移动支付平台百分百应用时客流量的情况，再带入第二问的盈利模型求出第三方支付平台的盈利情况。

4.4 问题四的分析

在前三问的基础上，我们根据盈利模型求得的结果，为第三方移动支付平台的推广效益做出预测以及建议。通过结果和数据分析得到的结果，对公司发展进行可行性意见报告，用以作为公司决策的辅助。并且在可能的情况下，给出盈利最大的情况，并且给出可行的方案。

五、问题一模型的建立与求解

5.1 数据预处理

为了分析该城市人群出行的特征，我们以乘车时间，支付方式，不同支付方式支付比重等不同的参数，来分析附件所给数据，希望能以此总结出出行支付特征。

由附件 1 表格给出的数据，我们需要找出乘车支付方式与乘车时间的关系，以此来分析乘车人的支付特征。

STEP1：读入数据——利用 PANDAS 读入给出表格中的数据。

STEP2：列名修改——由于所给表格中列名不同，为了后续的数据合并正常，我们将列名修改为统一列名。

STEP3：合并数据——将 28 个表格中的数据合并，进行整理。

STEP4：去掉异常数据——为了分析移动支付与公交卡支付两种支付方式，我们将所有显示 null（没有刷卡）和 0001-1-1（没有刷卡记录，原因是机器故障，但是仍然记录刷卡方式。）的数据不计入计算和输出图表中。

5.2 数据分析

由数据预处理阶段可知，我们将前期工作做完之后，利用 PANDAS 输出可视化图表，并根据直观的图表进行数据分析。

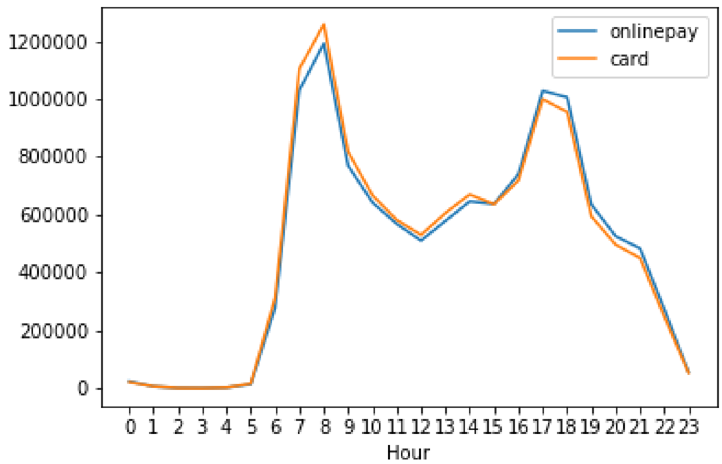


图 5-1 两种支付方式消费次数随小时变动图

1. 将一天划分为 24 小时作为横坐标，将使用两种方式的支付次数分别累和，设置图中橘黄色线条代表利用公交卡上车的支付次数，蓝色线条表示利用第三方移动支付平台的支付次数。由此可以得到图表如图 5-1 所示。

23 点至第二天 5 点的这一段公交车与地铁不运行的时间，图中数据表明，在 5 点至 7 点这段时间内，橘黄色线条与蓝色线条几乎重合，这意味着：利用公交卡和第三方移动平台支付的支付次数几乎相等，在 7 点至 15 点这段时间，橘黄色线条略高于蓝色线条，其中，7 点至 8 点这段时间橘黄色线条明显高于蓝色线条，即公交卡支付次数明显

多于第三方支付平台的支付次数。15 点至 22 点期间则是蓝色线条略高于橘黄色线条，又以 17 点至 18 点这段时间蓝色线条明显高于橘黄色线条，即第三方移动支付平台支付次数明显高于公交卡支付次数。

其原因可能是因为早高峰时有较多不熟悉使用第三方支付平台的人（例如年龄较大的长者）会选择在这个时段出行，所以造成使用公交卡的支付次数会较多。而晚高峰大多是由下班和下学的年轻人组成，而他们对第三方支付平台的依赖性更高，所以才会造成图示的线条。

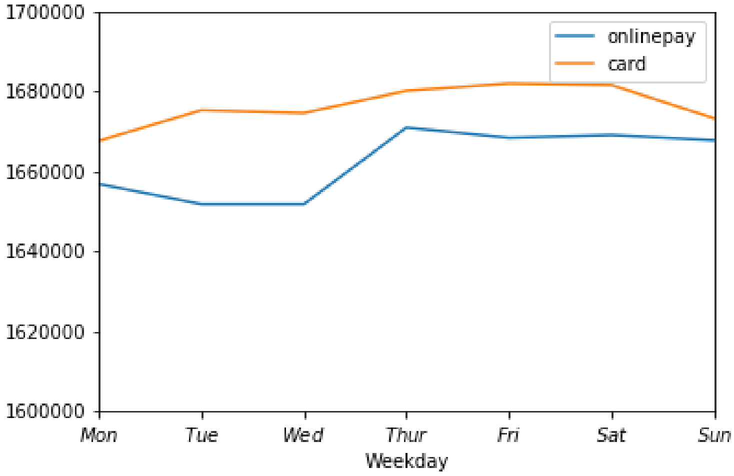


图 5-2 两种支付方式消费次数随星期天数变动图

2. 将所给的 28 天的数据按星期划分为周一至周日七天，并且仍然将支付次数累和，设置图中橘黄色线条代表利用公交卡上车的支付次数，蓝色线条表示利用第三方移动支付平台的支付次数，由此可以得到的折线图如图 5-2 所示。

   由折线图可以看出使用公交卡和第三方移动支付平台的人数差别不大（这里需要注意的是我们为了可以看清楚折线之间的具体情况，将整个纵轴的取值间隔减小，但是实际上两者数字差别仍然不大。），但是总体来说，使用公交卡的支付次数是大于使用第三方移动支付平台的支付次数。并且能够看出，两者在周三时差别最大，在周四和周日时差别最小。

   其原因除去统计数据时的一些误差外，可能与人们出行选择的偏好有关。在工作日出行的人群会在周四达到峰值，并且之后会趋于稳定，并且在周末时开始减少。而由于第三方支付平台的推广力度可能并没有非常普及，所以能够看出来使用公交卡支付的支付次数还是略高于移动支付的支付次数，但是两者差别非常小。

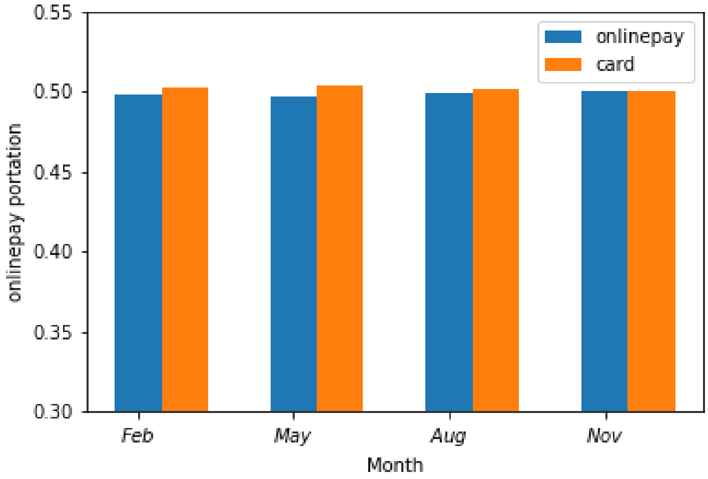


图 5-3 两种支付方式消费次数月份占比条形图

3. 附件 1 所给的数据分别是 2 月、5 月、8 月、11 月的各七天收集到的数据，按月份排列横坐标，将公交卡支付和第三方移动支付平台支付所占比例作为横坐标，求出的图表如图 5-3 所示。

　　由图中我们可以看出两者差距非常小，也就是说使用公交卡的支付次数和使用第三方支付的支付次数基本相同。我们可以看出在 2 月、5 月以及 8 月公交卡的支付次数都相较第三方支付平台的支付次数的占比要少。而 11 月两者达到平衡。并且可以看出来，随着月份的增加，第三方支付移动平台的支付次数也在小幅增长。

　　其主要原因是因为全国各个城市大概是从 2017 年开始推广使用移动支付平台，到 2017 年年底推广工作基本完成，所以随着月份的推移，使用第三方支付平台的支付次数也会相应增长，反映于数据图表上就是可以看到第三方支付平台的使用占比在缓慢的上升。

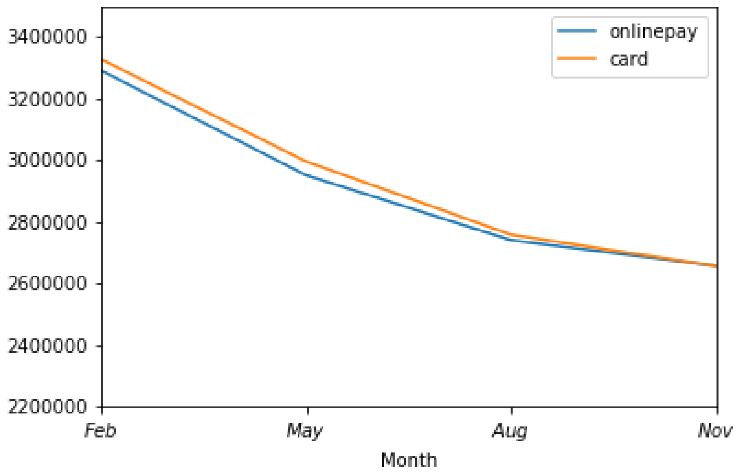


图 5-4 两种支付方式消费次数随月份变动图

4. 根据附录 1 给出的数据，我们以四个月份为横坐标，将使用两种方式的支付次数累和，设置图中橘黄色线条代表利用公交卡上车的支付次数，蓝色线条表示利用第三方移动支

付平台的支付次数，得到的两条曲线如图 5-4 所示。

由图中曲线所示，可以看到随着月份的增加，总体使用两种方式的支付次数都在减小，并且减小的幅度几乎相同。同时，在减小的过程中，使用公交卡的支付次数始终高于使用第三方移动支付平台的支付次数，直到 8 月份以及 11 月份两者几乎相等。

其原因可以推测几点：一.由于统计问题，后三个月份本来采集的样本数量就少于前几个月。二.由于我们在一开始就排除了未刷卡以及刷卡不成功的选项，所以还有可能是随着月份的推移，刷卡不成功的次数越来越高，则统计数据的样本会越来越少。三.随着月份的推移，选择公交卡乘车和第三方移动支付平台支付的人都逐渐减少，其原因可能是选择乘车的人减少，也可能是选择现金支付的次数增加。

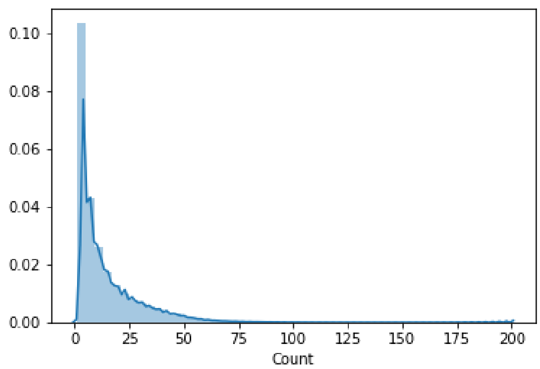


图 5-5 用户 28 天出行次数和分布图

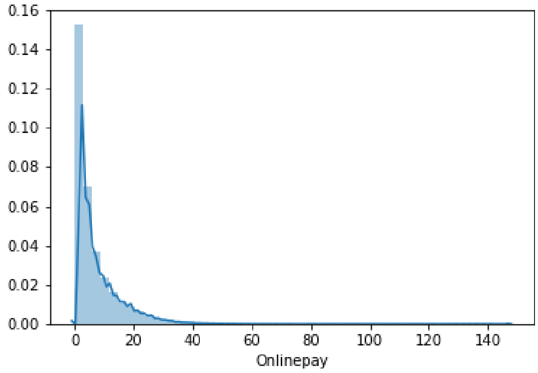


图 5-6 用户 28 天出行移动支付次数和分布图

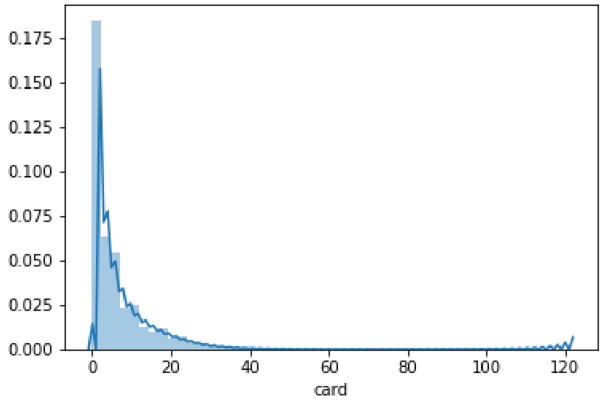


图 5-7 用户 28 天公交卡支付出行次数和分布图

5. 根据附录 1 给出的数据，我们将给出的 28 天为样本，以出行次数 200 次为界，将在 28 天内出行超过 200 次的数据视为异常数据，排除掉这一部分异常数据，以出行次数为横坐标，该次数所占所有人群出行次数的比例为纵坐标，得出直方图如图 5-5、5-6、5-7 所示。其中，5-5 为所有人群的两种方式总和的出行次数比例，5-6 为使用第三方支付平台的人群出行次数比例，5-7 为使用公交卡的人群出行次数比例。

由图 5-5 中数据我们可以看到，绝大多数人群在取样的 28 天内出行次数分布在 0 次-25 次的人数最多，最高峰可取到占所有人的 10%左右。由图 5-6 中数据我们可以看到，绝大多数使用第三方支付平台的人群在取样的 28 天内出行次数分布在 0 次-20 次的

人数最多，最高峰可取到占所有人的 15%左右。由图 5-7 中数据我们可以看到，绝大多数使用公交卡的人群在取样的 28 天内出行次数分布在 0 次-20 次的人数最多，最高峰可取到占所有人的 18%左右。

由此可以推断，人们在取样的 28 天内一般出行次数是在 0-20 次左右，同时，在 0-20 次出行的人群中，使用公交卡出行的人群多于使用第三方支付平台的人，也就是说在低频率出行的人中，更多的人选择了公交卡出行。

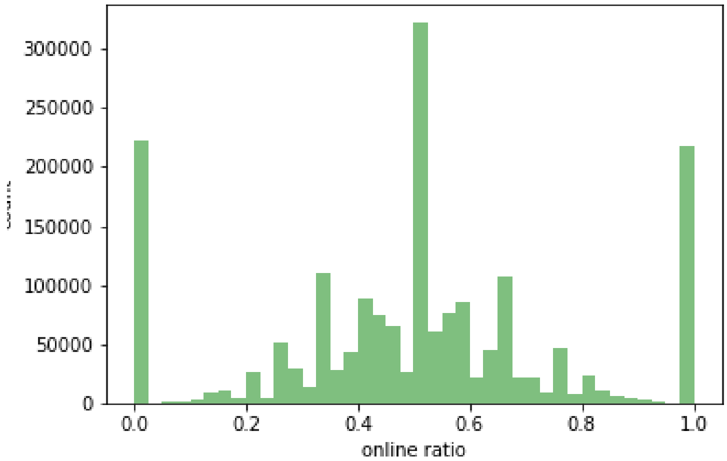


图 5-8 用户移动支付次数与出行总次数比值分布图

6. 根据附录 1 的数据，我们将每个人的数据分开，并且将每个人使用第三方移动支付平台的支付次数除以这个人两种方式支付的总次数作为横坐标，将和累计作为纵坐标，得到的分布图如图 5-8 所示。

由图中我们可以看出比值在 0.5 的人数最多，即有大多数人选择第三方移动支付平台和公交卡支付的概率相同。同时，在 0 和 1 的人数也较多，即单一的选择公交卡出行或选择第三方支付平台出行的人较多。

在 0 和 1 分布这么多人的原因是由于有很多人只记录了其一次的出行，当乘客这一次出行就选择了公交卡出行的话，他就被分布在了 0 的记录点上，反之，如果这次选择的是第三方支付平台支付的话就会被记录到 1 的记录点上。

为了减小这类一次出行的人对于结果的影响，我们再次筛选，将样本记录的 28 天内所有出行次数在 50 次以下的人全部排除掉，得到的分布图如图 5-9 所示。

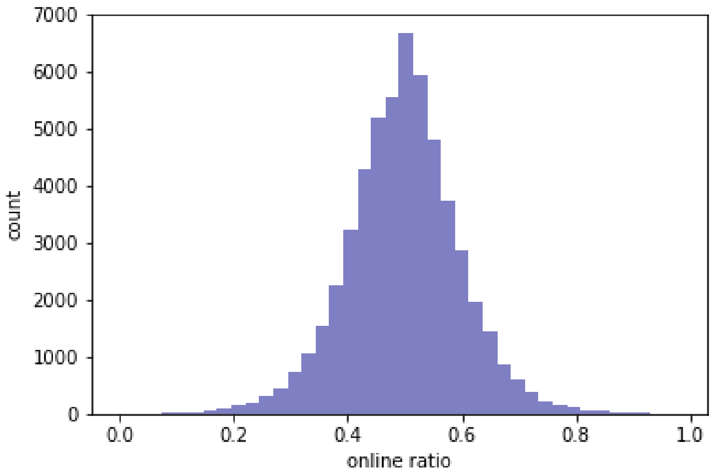
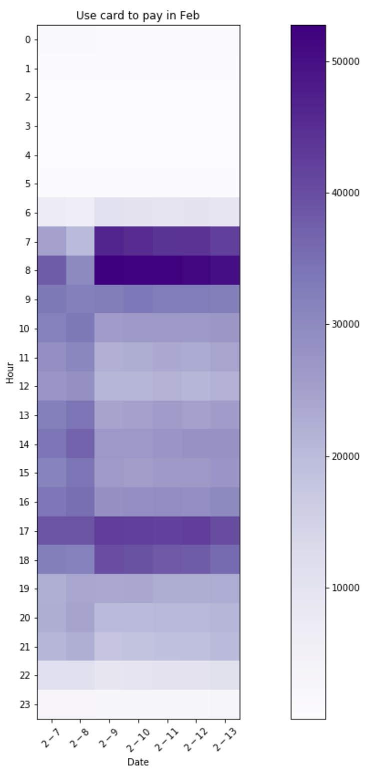


图 5-9 高频用户移动支付次数与出行总次数比值分布图

在去掉出行次数较小的人群后，我们可以看到数据分布明显集中于 0.4-0.6 的区间，在 0.5 处取得峰值，基本呈现正态分布，即大多数人是选择第三方移动支付平台和公交卡出行的概率相同，或者基本持平。由此我们可以看出，在高频出行的人群中，绝大多数人会既选择公交卡出行也选择第三方移动支付平台出行。



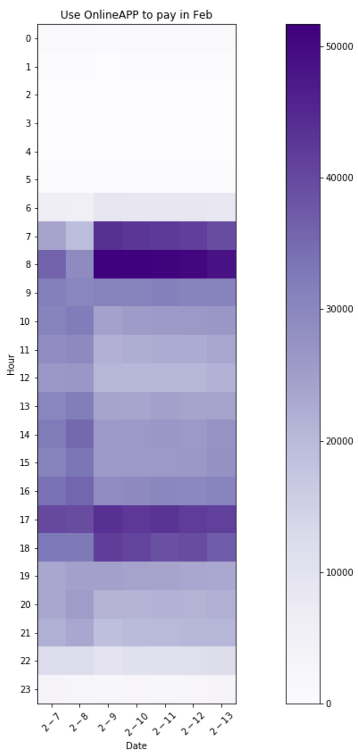


图 5-10 二月公交卡支付次数热图　　图 5-11 二月移动支付次数热图

7. 根据附录 1 给出的数据，为了能够直观的看出在每一天的不同时段使用不同支付方式的支付次数的聚集程度，我们设置横坐标为采集数据的每一天（以二月份为例），纵坐标

设置为一天内的 24 小时，利用颜色深浅来表示在该段时间支付上车的人数累和，颜色越深代表支付次数越高。根据数据生成的热图如图 5-10 和图 5-11 所示，其中 5-10 表示使用公交卡支付的支付次数，图 5-11 表示使用第三方移动支付平台支付的支付次数。

根据图表我们可以看出，两个图表中共同存在的一些特点如下：在一天之内，明显存在两个高峰时段，分别处于早晨 8 点和晚上 17 点，支付次数明显升高。并且在收集数据的这一周内，2 月 7 日和 2 月 8 日这两天的支付次数明显少于 2 月 9 日至 2 月 13 日这五天，同时，2 月 8 日在 14 点时也出现了一个小高峰。同时也可以看到公交卡支付和第三方支付平台支付的支付次数基本差不多。

其中原因存在于每一天的高峰时段，都集中在早晨 8 点和晚上 17 点左右，其乘客主要是大多数的下班下学的年轻人。在 7 日和 8 日出现的支付次数减少，可能是统计误差，也可能是这两天乘坐公共交通的人数减少。

5.3 数据分析小结

根据附件 1、2 给出的数据，我们进行了多方面的分析。

首先，我们能够看出，采样取得的数据时间非常科学，选择 2、5、8、11 这几个月都避开了寒暑假这种高峰时段，以及在这几个月采集的数据中，也同样避开了带有放假的节假日期间，选择的都是具有普遍意义的七天。采集七天的数据也可以很好的看出在一周的时间内，支付次数的变化情况，利于反映现实情况。

其次，在数据分析之前，我们经过讨论决定将异常数据去掉，比如支付方式为不刷卡的支付方式，以及既不采用公交卡支付也不采用移动支付方式的其他方式，还有刷卡不成功的支付次数，减除掉这一部分异常数据后我们利用软件对数据进行了分析。

再次，我们通过分析一天内 24 小时的支付次数变化情况，能够看出一天中的支付高峰，以及人们对于不同支付方式的倾向选择。 之后我们分析了一周内、不同月份的支付次数变化情况以及不同月份之间人们对两种支付方式的倾向性。最后我们将不同的人群分类，分别分析不同出行次数的人群对于两种支付方式选择的倾向性。

最后，我们得到了如下几个结论：

1) 在一天的时间内，在早晨 7-9 点会出现早高峰以及晚 17-19 点会出现晚高峰，支付次数达到一天峰值，并且在早高峰选择公交卡支付的次数多于移动支付次数，晚高峰则情况相反。
2) 在一周的时间内，数据反映出的结果为使用公交卡支付的支付次数略微高于使用第三方支付平台支付的次数，并且在周日时会出现使用两者的次数差别最小，在周三差别最大。
3) 在不同的月份里，随着月份的推移，公交卡的支付次数略高于移动支付次数，且与第三方移动支付平台的支付次数逐渐靠近，同时两者都在减少。
4) 低频出行的人群和高频出行的人群对于公交卡支付和第三方移动支付平台支付的倾向选择性大致相同，即人们在面对选择支付方式的人，约有一半的人选择

公交卡支付，而另一半的人选择第三方移动平台支付。

5) 对于每一个乘客而言，在他出行的许多次中，排除掉出行次数过低的人群，我们分析数据得到的结果是绝大多数人们会选择同时使用两种支付方式，并没有显现出来明显的对于某种支付方式的倾向性。

所以，根据附件 1 的数据我们可以得到以上的一些该城市的人出行的支付特征的初步结论，并且分析了可能的原因。

六、问题二模型的建立与求解

6.1 第三方支付平台的盈利模式

随着第三方支付平台的不断发展和移动信息技术的进步，第三方支付已经成为了具有良好发展前景的行业。运营模式和盈利模式的多样化使得其具有了强大的市场竞争力。在正常的第三方支付平台上有如下四种盈利模式：

1) 手续费

手续费是第三方支付平台向用户收取相应的手续费与银行收取的手续费用之差。第三方提供的个人服务有转账、缴费、提现以及短信的安全支付提醒和外币支付等；提供的企业服务则是查询金额、对账、追收尾款及退还钱款等清算交易相关服务的手续费。

手续费的收取比例区间大概在：$0.08\%\sim1.25\%$之间，转账和提现的收取比例为0.1%。不同服务类型的手续费不尽相同，同时，相同服务的不同金额的交易其手续费的收取比例也会有所不同。

2) 广告费

第三方支付平台作为使用频率较高的支付手段，由于其便利的移动客户端和相应的互联网平台，使它具有极强的宣传能力。因此第三方支付平台可以通过进行广告位收取商家的宣传广告费用作为盈利途径之一。

3) 沉淀资金的利息收入

沉淀资金的定义如下：支付机构在办理客户委托的支付业务时实际收到的预存货币代付资金。

根据《支付机构客户备付金管理办法》中的相关条例：第三方支付平台的风险准备金的比例不得低于备付金在银行账户利息所得的10%，则第三方支付平台的最高盈利则是利息的90%。根据相关条例和银行的规章制度，部分的备付金可以以中国人民银行批准的形式进行存放。其存放形式大致有：单位定期存款、活期存款、协定存款以及单位通知存款。目前来说，是以活期存款作为主要的存款方式，但是其期限不得超过三个月。此类存款的协议存款率大致在$4\%\sim5\%$，同时每笔存款还应向银行支付0.78%的手续费。

4) 服务费

第三方支付平台在提供基础的支付系统的基础上还能有支付解决方案和各类增值

服务。随着，平台的不断扩大，城市生活服务与其不断结合，提供了多种缴费的途径，如电费、水费、手机话费等不同方面。平台在提供给个人或者组织更为便利的支付途径时，增加了在市场上的竞争力，同时，收取部分服务费用作为平台的盈利模式之一。随着技术和服务能力的不断提升，平台具有更多的竞争力、吸引客户扩大影响。

6.2 第三方支付平台的支出模式

1) 广告费用

第三方支付平台与公共交通进行合作，通过平台进行交通费用的日常充值、使用等服务，增加新的支付模式，便利更多居民。为了宣传第三方支付平台的新的服务，使得更多的乘客使用第三方支付平台进行支付，需要在报纸、公交车的广告牌、车站的广告牌等处进行前期宣传，增加影响力。

2) 接入费用

当前第三方支付平台的接口接入模式主要有电脑网站支付、手机网站支付、APP 支付等，各第三方支付公司的支付接口费率也趋于相同，一般行业费率在0.6％左右，游戏、娱乐等虚拟业务的费率为1％。

3) 固定支出

第三方支付平台在运营阶段需要在人员管理、新项目服务、银行对接等方面进行固定的投入。如基础设施的建造、服务器的建立和管理以及和银行进行备付金的相关协议的协商。因此，需要进行部分一次性的固定支出用于基础设等、还需要进行部分的持续性支出如人员的劳务费等。

6.3 第三方支付平台在公共交通系统的盈利模式

公共交通系统具有明确的经济特性，它是满足人们在公共区域里面的空间上的位移。公共交通既要满足民生要求也要从服务对象处获得报酬，是在实现自身的经济效益的时候也会对周围的区域产生经济的影响。同时它具有流量大、金额多、不便于管理等特性。除此以外，公共交通系统的运营受到外界的影响因素很大，政策以及线路的规划都会对其产生影响。

根据上述的盈利模式和支出模式结合公共交通系统的环境与特性可以得出其盈利的专属模式：

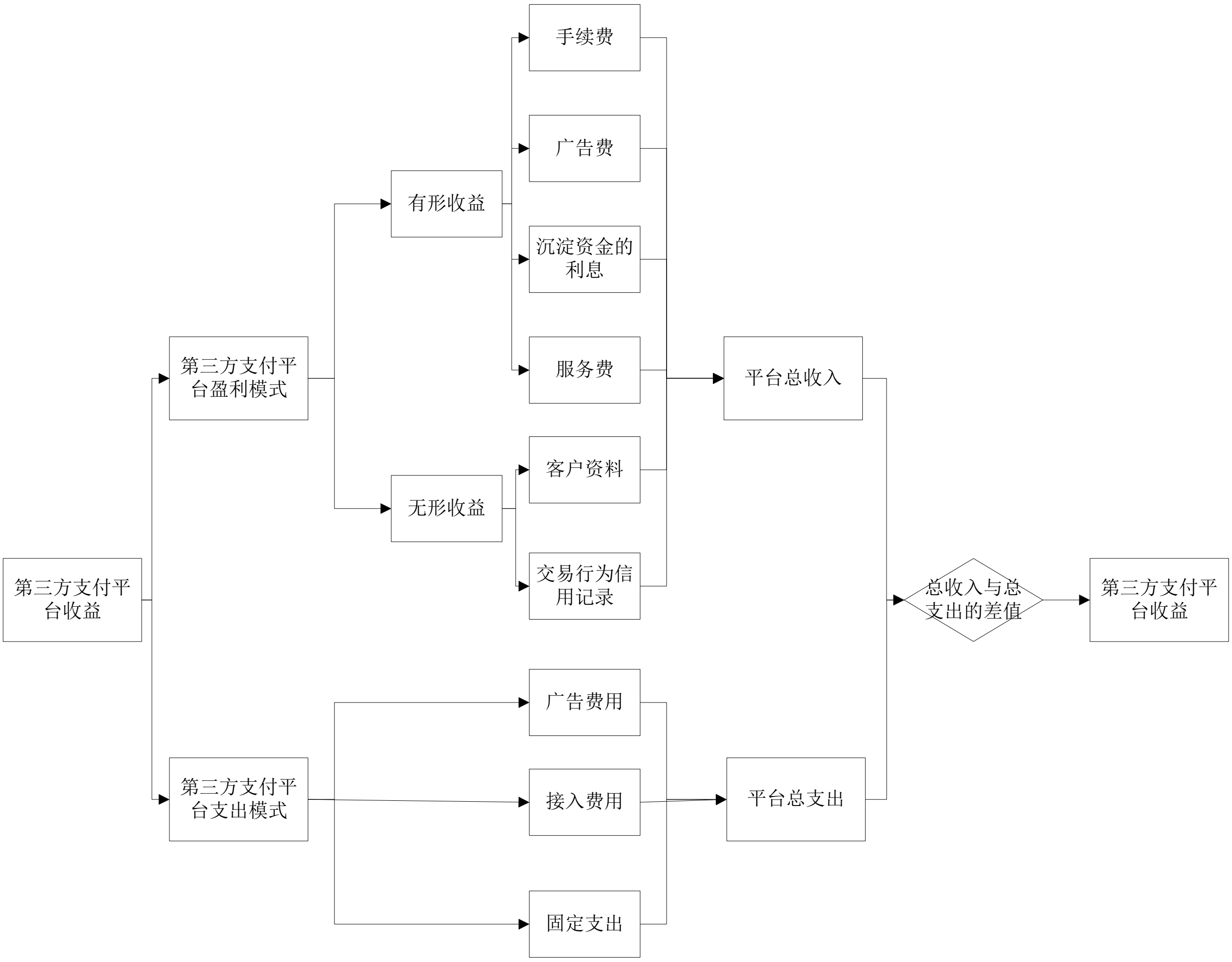


图 6-1 盈利模式示意图

6.4 第三方支付平台的商业盈利数学模型的建立

根据公共交通的支付特点结合第三方支付平台的盈利模式，我们不难得出上述的收入-支出的流程图。则有：

$$W_0 = I_0 - O_0 \tag{6-4-1}$$

其中W_0为模型的盈利、I_0为模型的总收入、O_0为模型的总支出。

由上述的流程图可知总收入I_0的数学表达式为：

$$I_0 = I_1 + I_2 + I_3 + I_4 \tag{6-4-2}$$

其中I_1为沉淀资金的利息收入、I_2为手续费用的收入、I_3为服务费用的收入、I_4为广告费用的收入。

沉淀资金的利息I_1的数学计算：沉淀资金的利息生成属于动态变化中的线性规划问题，其线性动力系统公式为：

$$I_{n+1} = r\left(I_n + X_n + X_{n-1} + X_{n-2}\right) \quad （r 为常数） \tag{6-4-3}$$

其中X_i为第i个月的充值总金额即沉淀资金（$i = 1 \ldots n$）

增长系数r的确定：

$$r = a_1 \bullet a_2 \bullet (1 - a_3) \tag{6-4-4}$$

其中a_1为协议存款率、a_2为沉淀资金利息的最高获得率、a_3为与银行协议的手续费。

根据银行的相关协议，活期存款作为主要的存款方式，其存款期限不得超过三个月。则有如下的计算公式：

$$\begin{aligned} I_{n+1} &= \left(r^n + r^{n-1} + r^{n-2}\right) \bullet X_1 + \left(r^{n-1} + r^{n-2} + r^{n-3}\right) X_2 + \ldots\ldots \\ &+ \left(r^4 + r^3 + r^2\right) X_{n-3} + \left(r^3 + r^2 + r\right) X_{n-2} + \left(r^2 + r\right) X_{n-1} + r X_n \end{aligned} \tag{6-4-5}$$

其中$n \geq 1$、X_i为每个月的充值金额即沉淀资金（$i = 1....n$）

手续费用I_2的数学计算：手续费的收取比例是与其充值金额有关，则此时有关于I_2的数学公式为区间函数，其间隔点的选取与协商的政策相关。

$$I_2 = \begin{cases} r_s \bullet X_{si} \bullet n & r_s = 0.08\% \quad X_{si} \leq 100 \\ r_{s1} \bullet X_{si} \bullet n & r_s = 0.1\% \quad X_{si} > 100 \end{cases} \tag{6-4-6}$$

其中r_s、r_{s1}为手续费的收取比例、X_{si}为第i月的单次充值金额、n为充值的人数。

服务费用I_3的数学计算：服务费用的收取比例与其金额有关，且具有费用的下限和上限的限制。

$$I_3 = r_f \bullet X_{fi} \bullet n, \quad r_f \bullet X_{fi} \subseteq [0.5, 10] \tag{6-4-7}$$

其中r_f为服务费的收取比例、X_{fi}为第i个月的单次金额、服务费用$r_f \bullet X_{fi}$的区间为$[0.5, 10]$、n为充值的人数。

广告费用I_4的数学计算：广告费的金额与人数有关，在低区间内广告费用为常数、在高区间内广告费用与人数呈现出正比的函数关系。

$$\begin{aligned} I_4 &= A_g & n \leq N \\ I_4 &= A_g + n \bullet X_g & n > N \end{aligned} \tag{6-4-8}$$

其中A_g为在低区间内一次性支付的广告费用、X_g为在高区间内单个移动平台赚取的广告费、n为使用平台的人数、N为使用人数的阈值。

由上述流程图可知总支出O_0的数学表达式为：

$$O_0 = O_1 + O_2 + O_3 \tag{6-4-9}$$

其中O_1为前期投入的广告宣传费用、O_2为移动端接入的费用、O_3为固定的支出费用。

前期广告的宣传费用O_1的计算公式为：

$$O_1 = m \bullet X_c \tag{6-4-10}$$

其中m为前期投入宣传的次数、X_c为单次宣传的费用。

移动端接入费用O_2的计算公式为：

$$O_2 = r_j \bullet X_j \bullet n \tag{6-4-11}$$

其中r_j为接入费用的收取比例、X_j为费用的金额、n为接入的数目。

固定支出费用O_3的数学计算公式为：

$$O_3 = A_j + A_x + n \bullet c \bullet X_r \tag{6-4-12}$$

其中A_j为基础设施的投入成本、A_x为新项目的固定成本、c为员工个数、X_r为员工工资、n为工作的月数。

将上述收入与支出的数学计算公式进行整合可得：

$$W_0 = I_0 - O_0$$

$$W_0 = I_1 + I_2 + I_3 + I_4 - (O_1 + O_2 + O_3)$$

$$W_0 = r(I_n + X_n + X_{n-1} + X_{n-2}) + r_s \bullet X_{si} \bullet n + r_f \bullet X_{fi} \bullet n + A_g + n \bullet X_g -$$
$$(m \bullet X_c + r_j \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r)$$

其中有：

$$\begin{cases} n \geq 1, i = 1......n \\ r_s = 0.08\%, X_{si} \leq 100 \\ r_s = 0.1\%, X_{si} > 100 \\ r_f \bullet X_{fi} \subseteq [0.5,10] \\ n \leq N, X_g = 0 \\ r \geq 0, X \geq 0, r_s \geq 0, X_{si} \geq 0, r_f \geq 0, X_{fi} \geq 0, A_j \geq 0, A_x \geq 0 \end{cases} \tag{6-4-13}$$

6.5 定量分析盈利数学模型模型

6.5.1 总收益I_0的定量分析：

1) 沉淀资金I_1的定量分析计算：增长系数r的确定：

$$r = a_1 \bullet a_2 \bullet (1 - a_3) \tag{6-5-1}$$

其中a_1为协议存款率、a_2为沉淀资金利息的最高获得率、a_3为与银行协议的手续费。则有：

$$\begin{cases} a_1 = 4.5\% \\ a_2 = 90\% \\ a_3 = 0.78\% \end{cases} \tag{6-5-2}$$

代入得： $r = 0.045 \bullet 0.9 \bullet (1 - 0.0078) = 0.0401841 \approx 0.04$ (6-5-3)

其中X_i为第i个月的充值总金额即沉淀资金（$i = 1....n$），则X_i为自变量。则有：

$$I_{n+1} = r(I_n + X_n + X_{n-1} + X_{n-2}) = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) \tag{6-5-4}$$

2) 手续费用I_2的定量分析计算:手续费用的计算公式为:

$$I_2 = r_s \bullet X_{si} \bullet n \tag{6-5-5}$$

此函数为区间函数，间隔点由收费制度确定为 100，即$X_{si} = 100$时为间断点：

$$\begin{cases} r_s = 0.08\%, X_{si} \le 100 \\ r_s = 0.1\%, X_{si} > 100 \end{cases} \tag{6-5-6}$$

其中 X_{si} 为第 i 月的单次充值金额、n 为充值的人数，这两者构成了此函数的自变量。

3) 服务费用 I_3 的定量分析计算：服务费用的计算公式为：

$$I_3 = r_f \bullet X_{fi} \bullet n \tag{6-5-7}$$

根据相关的支付条例可得：

$$r_f = 0.1\% \tag{6-5-8}$$

其中有 X_{fi} 为第 i 个月的单次金额、n 为充值的人数，两者均为函数的自变量。但是由于相关条例的限制 I_3 有相应的取值范围即 $r_f \bullet X_{fi} \subseteq [0.5,10]$，带入得：

$$\begin{cases} X_{fi\min} = \dfrac{0.5}{0.1\%} = 500 \\ X_{fi\max} = \dfrac{10}{0.1\%} = 10000 \end{cases} \tag{6-5-9}$$

则有分段函数成立：

$$\begin{cases} I_3 = 0, X_{fi} < 500 \\ I_3 = r_f \bullet X_{fi}, 500 \le X_{fi} \le 10000 \\ I_3 = 10, X_{fi} > 10000 \end{cases} \tag{6-5-10}$$

4) 广告费用 I_4 的定量分析计算：

$$\begin{cases} I_4 = A_g & n \le N \\ I_4 = A_g + n \bullet X_g & n > N \end{cases} \tag{6-5-11}$$

关于阈值 N 的确定：决定 N 的影响因素有两个，一方面是使用移动平台的的用户人数 n，另一方面是影响力的隐函数 $f(n)$。

则关于阈值 N 的函数确定：

$$N = \varphi \bullet n \bullet f(n) + \sigma \tag{6-5-12}$$

其中 n 为第三方支付平台的移动端的人数、φ 是用户人数的预测系数、$f(n)$ 是影响力的隐函数、σ 是误差值。

总支出 O_0 定量分析

1) 前期广告宣传费用 O_1 的定量分析计算：

$$O_1 = m \bullet X_c \tag{6-5-13}$$

其中 m 为前期投入宣传的次数、X_c 为单次宣传的费用。m、X_c 均为函数的自变量且与第三方支付平台的签订的协议以及战略目标有关。根据分析数据、收集相关资料，取 m、X_c 为各类型的算数平均值：

$$\begin{cases} m = \dfrac{m_1 + m_2 + m_3 + ... + m_n}{n} = \dfrac{\sum_{i=1}^{n} m_i}{n} \\ X_c = \dfrac{X_{c1} + X_{c2} + X_{c3} + ... + X_{cn}}{n} = \dfrac{\sum_{i=1}^{n} X_{ci}}{n} \end{cases} \tag{6-5-14}$$

2) 移动端接入费用O_2的定量分析计算：

$$O_2 = r_j \bullet X_j \bullet n \tag{6-5-15}$$

其中接入费用的收取比例r_j和费用的金额X_j为主要的影响变量。根据第三方行业的支付准则：

表 6-1 第三方行业的支付准则表

产品名称	应用场景	接入费	费率	保证金
AAP 支付	移动应用	0 元	一般行业：0.6%	0 元
手机网站支付	移动网页	0 元	一般行业：0.6%	0 元
电脑网站支付	PC 网页	0 元	一般行业：0.6%	0 元
当面付	扫码	0 元	一般行业：0.6%	0 元
转到账户	企业付款	0 元	免费	0 元
支付宝批量付款	企业付款	0 元	单笔费率：0.5%	0 元

根据表格可知：接入费用的收取比例

$$r_j = 0.6\% 。\tag{6-5-16}$$

3) 固定支出费用O_3的定量分析计算：

$$O_3 = A_j + A_x + n \bullet c \bullet X_r \tag{6-5-17}$$

其中A_j为基础设施的投入成本、A_x为新项目的固定成本、c为员工个数、X_r为员工工资、n为工作的月数。A_j、A_x为确定的数值；c、X_r、n均为变量，与实际情况有关。

6.5.2 总盈利W_0的定量分析：

$$W_0 = I_0 - O_0 \tag{6-5-18}$$

$$W_0 = I_1 + I_2 + I_3 + I_4 - (O_1 + O_2 + O_3) \tag{6-5-19}$$

带入各个函数公式：

$$W_0 = r(I_n + X_n + X_{n-1} + X_{n-2}) + r_s \bullet X_{si} \bullet n + r_f \bullet X_{fi} \bullet n + A_g + n \bullet X_g - \\ (m \bullet X_c + r_j \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r) \tag{6-5-20}$$

将定量分析后的数值代入：

根据不同的约束分类，我们可以得到如下表达式及其约束条件：

①：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.08\% \bullet X_{si} \bullet n + 0.1\% \bullet X_{fi} \bullet n + A_g + n \bullet X_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} \le 100 \\ 500 \le X_{fi} \le 10000 \\ n > N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \tag{6-5-21}$$

②：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.1\% \bullet X_{si} \bullet n + 0.1\% \bullet X_{fi} \bullet n + A_g + n \bullet X_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} > 100 \\ 500 \le X_{fi} \le 10000 \\ n > N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \tag{6-5-22}$$

③：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.08\% \bullet X_{si} \bullet n + 0 + A_g + n \bullet X_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} \le 100 \\ X_{fi} < 500 \\ n > N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \tag{6-5-23}$$

④：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.1\% \bullet X_{si} \bullet n + 0 + A_g + n \bullet X_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} > 100 \\ X_{fi} < 500 \\ n > N \\ X \geq 0, r_s \geq 0, A_j \geq 0, A_x \geq 0 \end{cases} \tag{6-5-24}$$

⑤：

$$W_0 = 0.04\left(I_n + X_n + X_{n-1} + X_{n-2}\right) + 0.08\%_0 \bullet X_{si} \bullet n + 10 + A_g + n \bullet X_g$$
$$-\left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\%_0 \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r\right)$$

$$\begin{cases} X_{si} \leq 100 \\ X_{fi} > 10000 \\ n > N \\ X \geq 0, r_s \geq 0, A_j \geq 0, A_x \geq 0 \end{cases} \tag{6-5-25}$$

⑥：

$$W_0 = 0.04\left(I_n + X_n + X_{n-1} + X_{n-2}\right) + 0.1\%_0 \bullet X_{si} \bullet n + 10 + A_g + n \bullet X_g$$
$$-\left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\%_0 \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r\right)$$

$$\begin{cases} X_{si} > 100 \\ X_{fi} > 10000 \\ n > N \\ X \geq 0, r_s \geq 0, A_j \geq 0, A_x \geq 0 \end{cases} \tag{6-5-26}$$

⑦：

$$W_0 = 0.04\left(I_n + X_n + X_{n-1} + X_{n-2}\right) + 0.08\%_0 \bullet X_{si} \bullet n + 0.1\%_0 \bullet X_{fi} \bullet n + A_g$$
$$-\left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\%_0 \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r\right)$$

$$\begin{cases} X_{si} \leq 100 \\ 500 \leq X_{fi} \leq 10000 \\ n < N \\ X \geq 0, r_s \geq 0, A_j \geq 0, A_x \geq 0 \end{cases} \tag{6-5-27}$$

⑧：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.1\% \bullet X_{si} \bullet n + 0.1\% \bullet X_{fi} \bullet n + A_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} > 100 \\ 500 \le X_{fi} \le 10000 \\ n < N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \qquad (6\text{-}5\text{-}28)$$

⑨：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.08\% \bullet X_{si} \bullet n + 0 + A_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} \le 100 \\ X_{fi} < 500 \\ n < N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \qquad (6\text{-}5\text{-}29)$$

⑩：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.1\% \bullet X_{si} \bullet n + 0 + A_g - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} > 100 \\ X_{fi} < 500 \\ n < N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \qquad (6\text{-}5\text{-}30)$$

⑪：

$$W_0 = 0.04(I_n + X_n + X_{n-1} + X_{n-2}) + 0.08\% \bullet X_{si} \bullet n + 10 + A_g + - \left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r \right)$$

$$\begin{cases} X_{si} \le 100 \\ X_{fi} > 10000 \\ n < N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \tag{6-5-31}$$

⑫：

$$W_0 = 0.04\left(I_n + X_n + X_{n-1} + X_{n-2}\right) + 0.1\% \bullet X_{si} \bullet n + 10 + A_g$$
$$-\left(\frac{\sum_{i=1}^{n} m_i}{n} \bullet \frac{\sum_{i=}^{n} X_{ci}}{n} + 0.6\% \bullet X_j \bullet n + A_j + A_x + n \bullet c \bullet X_r\right)$$

$$\begin{cases} X_{si} > 100 \\ X_{fi} > 10000 \\ n < N \\ X \ge 0, r_s \ge 0, A_j \ge 0, A_x \ge 0 \end{cases} \tag{6-5-32}$$

6.5.3 数学模型的精确计算：

以模型⑨为例进行计算：因为已知数据的不完善，在将已有系数进行量化分析后，根据附件提供的基础数据对未知的自变量进行预测分析。同时，结合相关的公共交通数据，使用数学归纳法对其进行合理的推断，从而得出以下数据：

表 6-2 模型取值表

I_n	X_n	X_{n-1}	X_{n-1}	X_{si}	X_{fi}	A_g	X_g	m_i	X_{ci}	X_j	n
8.1×10^{6}	6.6×10^{7}	6.6×10^{7}	6.6×10^{7}	20	20	1×10^{6}	15	3	1×10^{6}	20	3.3×10^{6}

则代入数学模型的公式进行计算可以得：

$$W_0 = I_0 - O_0 = (810 + 5.28 + 100 + 0 - 600 - 39.6 - 85)\times10^4 = 1906800 \tag{6-5-33}$$

综上所述，在正常的盈利模式下，基于已有的附件数据该第三方支付平台的盈利大致在1.9×10^6元左右，属于高盈利的第三方支付平台。

6.6 第三方支付平台盈利数学模型的分析

盈利模型的构成基于“盈利”的含义，即“盈利的值为总收入与总支出的差值”。表现为数学公式：$W_0 = I_0 - O_0$（W_0为模型的盈利、I_0为模型的总收入、O_0为模型的总支出）。在本问题中，需要先明确收入和支出的组成部分，然后根据各个部分的不同情况建立相应的数学模型。因为，第三方支付平台的经济特性其收入都具备“按比例收取相应费用”的特征，则在数学模型的建立过程中着重考虑了分段函数和线性规划这两个方面。对于关键变量的间隔点，即阈值的确定，通过建立复合函数甚至是隐函数来进行求解，旨在得出更为合理的分割方案和区间。在模型的建立过程中，很多比例系数r_i的

选择都参考现有的规章制度，尽量贴合真实数据。在最终得出的模型之中，仍有需要改进的部分，如：在确定广告费用I_4时，其阈值$N=\varphi \bullet n \bullet f(n)+\sigma$,其中影响力的隐函数$f(n)$和误差修正值$\sigma$都需要进行进一步的验证与计算，从而可以得到更为精准的定量计算结果。

七、问题三模型的建立与求解

7.1 整体分析

由题意，数据为$1/4$的公交和地铁安装移动支付设备后试运营阶段取得的数据，由于数据中无法区分公交出行与地铁出行时所采用的支付方式，因此，此问中不再考虑地铁与公交的具体差异。若需要估计该城市全部公交，实现公交第三方平台支付后的盈利情况，需按下述流程进行：

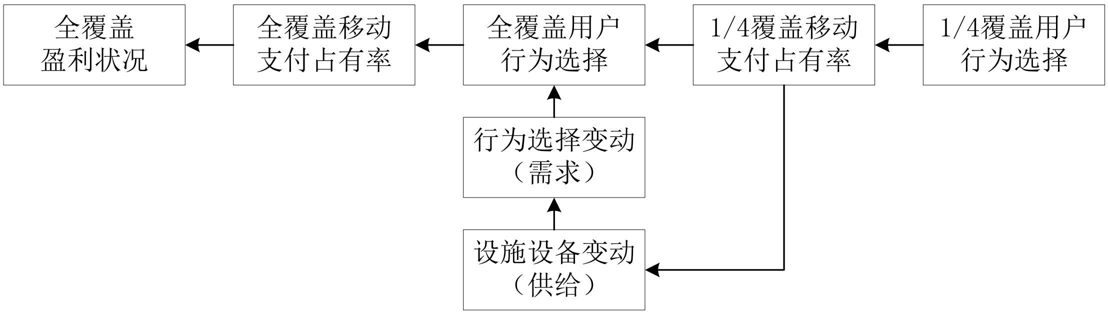


图 7-1 问题三分析流程图

若希望得知全覆盖条件下的盈利状况，则需要获知全覆盖条件下移动支付的占有率情况。移动支付占有率与用户的行为选择有着密切联系，用户行为选择是需求的体现形式之一。覆盖率的变化通过设施设备的变动体现，这是系统的供给部分，供给的变动会影响需求。因此我们需要通过现有问题一$1/4$覆盖率条件下的移动支付占有率及占有情况数据，建立模型刻画行为选择变动，从而预测全覆盖条件下的移动支付占有率，进而通过问题二建立的模型求得全覆盖条件下的盈利状况。

7.2 问题一占有率现状分析

由问题一数据分析结果可以看出，在移动支付设备$1/4$的情况下，移动支付次数与公交卡支付次数比值约为 1，即不考虑其他支付方式的情况下，移动支付约占比 50%。因此，当仅考虑已安装设备的车辆时，在这些车辆中，移动支付占比一定大于 50%。若需确定在已安装设备上移动支付设备的占有率具体为何值，则需要估计$1/4$覆盖车客流占总客流量的比例。由假设安装了移动支付设备的四分之一公交、地铁为该城市最热门的线路，以郑州市为例，将公交线路客流量由小到大排序，绘制线路日客流量分布曲线。经修正，曲线如下图所示：

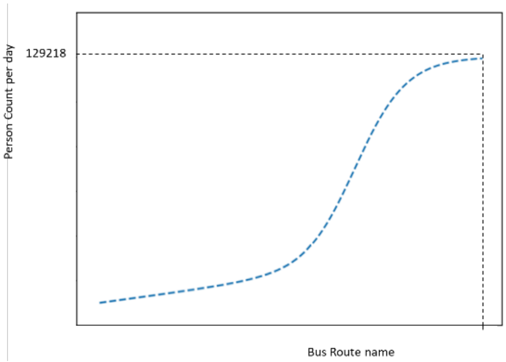


图 7-2 公交客流量累计分布图

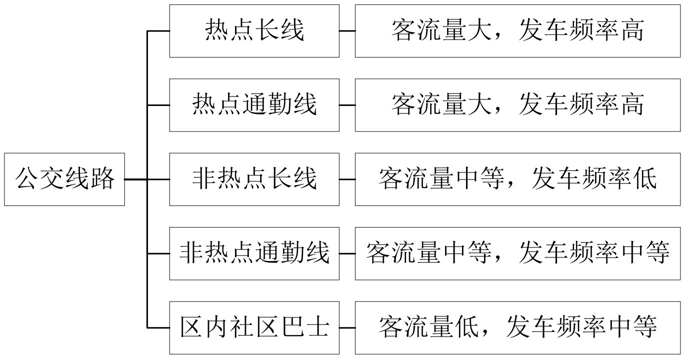


图 7-3 公交分类示意图

上图中横轴代表城市内的公交线路，纵轴为其对应的日平均客流总量，经计算可得，客流量排名前 1/4 的公交客流所占总客流的比例约为 74.86%，此处按照 75%计算。这主要是由于公交线路依照覆盖面与热门性可分为：热点长线、热点通勤线、非热点长线、非热点通勤线与区内社区巴士五种类型，公交类型不同，分担客流量与发车频率不同，但线路分布的广度不可缺少，造成了少量线路分担大量客流的状况。

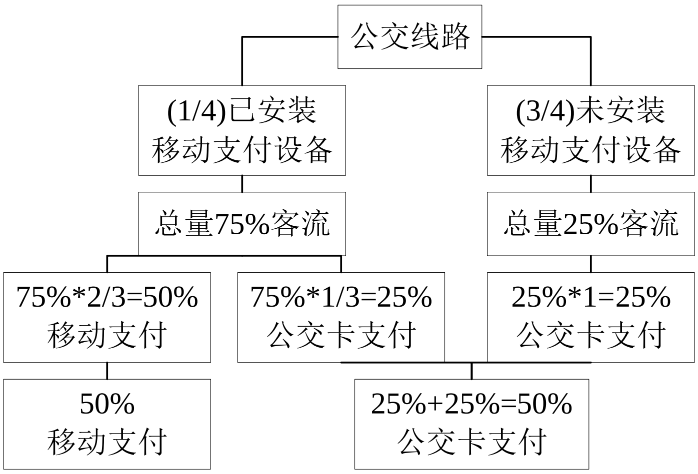


图 7-4 移动支付-公交卡支付比率图

由以上分析可知，在$^{1}/_{4}$安装了移动支付设备的公交中，移动支付占比为 66.6%，公交卡支付占比为 33.3%。

由于供给-需求关系的影响，以上占比会产生变化，供给-需求的变动关系如下图所示：

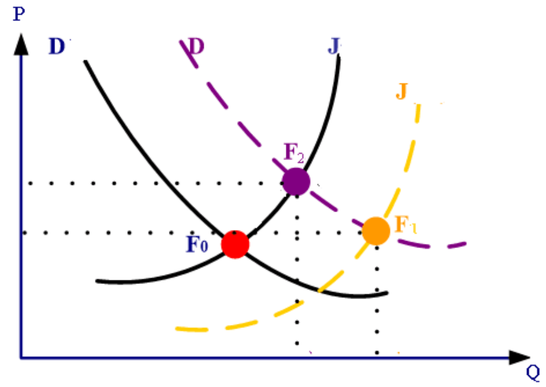


图 7-5 供求曲线变动示意图

初始状态，供给与需求如图中黑色曲线所示，F_0为平衡状态；随着供给的增加（黑色 J 曲线右移至黄色虚线），需求也产生变动（黑色 D 曲线右移至紫色曲线）。因此，接下来，采用集计与非集计两种角度分析设备全覆盖后，用户的选择情况。

7.3 集计方法——创新扩散模型

创新扩散模型是以创新扩散理论为基础建立的，由 E.M.Roger 提出。创新被定义为：一种被个人或其他采纳单位视为新颖的观念、时间或事物。[1]且创新具备相对的便利性、兼容性、复杂性、可靠性和可感知性五个要素。创新的采用者可以分为革新者、早期采用者、早期追随者、晚期追随者和落后者五种，创新的扩散模式如下图所示：

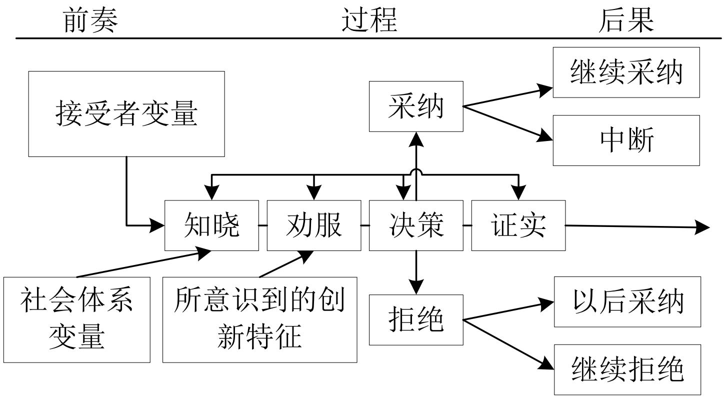


图 7-6 E.M.Roger 创新扩散模型

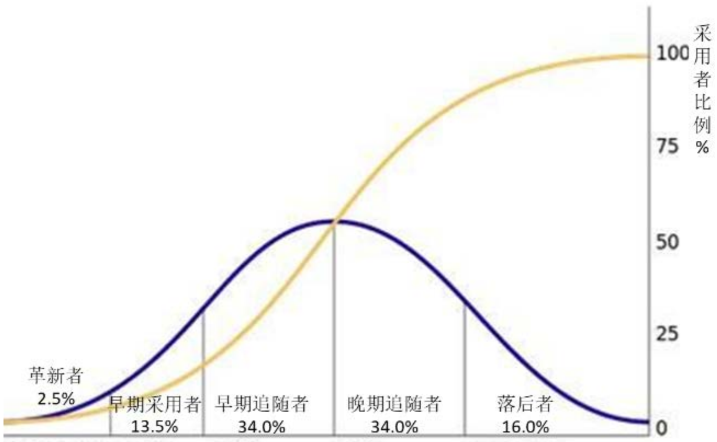


图 7-7 E.M.Roger 创新扩散模型

上述曲线说明，创新扩散过程中，约有 2.5%的人属于革新者，13.5%的人属于早期采用者，34%的人属于早期追随者，34%的人属于早期追随者，16%的人属于落后者。

移动支付的设备安装与推广过程本身属于创新的扩散过程，以该集计模型预估全覆盖状态下，采用移动支付人的比率，由于预估阶段为设备全覆盖后一段时间内的比率，因此此处不考虑落后者。预估全覆盖条件下，有 84%的人采用移动支付手段。

7.4 非集计方法——离散选择模型

离散选择模型属于多重变量分析的方法之一，是社会学、生物统计学、数量心理学、市场营销等统计实证分析的常用方法。该包含：决策者、有限个决策选项、选项的属性、

决策者的偏好四个要素。[2]离散选择模型基于效用理论，认为：①决策者选择某个选项会得到一定的效用②假设决策者理性，总是选择效用最大的选项③某个选项的效用由该选项的属性和决策者的偏好共同决定④效用具有不确定性，这种不确定源于对决策者偏好的认识不完全准确、对选项属性的认识不完全准确。

针对本问题，分析本问题的决策者、决策选项、选项属性如下：

- 人群：中青年、老年人
- 决策选项：移动支付、公交卡支付
- 选项属性：实际花费、取用时间、设备覆盖、便携程度、功能广泛性、使用难度

首先对于中青年人群进行分析，假设各选项的效用为属性的加权线性组合，有

$$u_1 = \widehat{u_1} + e_1 = \beta_6 x_6^1 + \beta_5 x_5^1 + \beta_4 x_4^1 + \beta_3 x_3^1 + \beta_2 x_2^1 + \beta_1 x_1^1 + \beta_0 + e_1 \quad (7\text{-}4\text{-}1)$$

$$u_2 = \widehat{u_2} + e_2 = \beta_6 x_6^2 + \beta_5 x_5^2 + \beta_4 x_4^2 + \beta_3 x_3^2 + \beta_2 x_2^2 + \beta_1 x_1^2 + \beta_0 + e_2 \quad (7\text{-}4\text{-}2)$$

其中u_1，u_2分别表示中青年人群选择移动支付和公交卡支付的效用；x 是属性，下标{1,2,3,4,5,6}对应 6 个属性，上标{1,2}对应移动支付与公交卡支付两种方式；β是偏好参数，e是随机变量，代表效用的不确定性。

由效用理论中的理论假设可知，决策者选择移动支付的条件是，移动支付的效用大于公交卡支付的效用：

$$u_1 > u_2 \quad (7\text{-}4\text{-}3)$$

$$\widehat{u_1} + e_1 > \widehat{u_2} + e_2 \quad (7\text{-}4\text{-}4)$$

整理得：

$$e_1 - e_2 > \widehat{u_2} - \widehat{u_1} \quad (7\text{-}4\text{-}5)$$

$$P_1 = P(u_1 > u_2) = P(e_1 - e_2 > \widehat{u_2} - \widehat{u_1}) \quad (7\text{-}4\text{-}6)$$

假设e为 logistic 函数，则有：

$$P_1 = \frac{e^{\widehat{u_1}}}{e^{\widehat{u_1}} + e^{\widehat{u_2}}} \quad (7\text{-}4\text{-}7)$$

$$P_2 = \frac{e^{\widehat{u_2}}}{e^{\widehat{u_1}} + e^{\widehat{u_2}}} \quad (7\text{-}4\text{-}8)$$

在设备覆盖为$1/4$的条件下，由前文分析知$P_1 = 66.6\%$，$P_2 = 33.3\%$，对于该条件下的x_1至x_6值进行量化，进而进行标准化，标准化公式为：

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7\text{-}4\text{-}9)$$

利用调查法，初步计算针对中青年与老年的偏好参数（偏好参数间比值），将标定好的 x 值、β值、P值带入 P 值计算公式，进一步去调整偏好参数，使之满足等式条件。

设备覆盖由$1/4$扩展到 1 的过程中，实际花费、取用时间、便携程度、功能广泛性、使用难度x_1, x_2, x_4, x_5, x_6属性不发生变化，设备覆盖x_3发生变化。利用离散选择模型，计算扩展后中青年、老年选择两种支付方式的概率。

表 7-1 全覆盖条件下选择概率

决策选项	人群	选择概率
移动支付	中青年	92.17%
	老年	25.88%
公交卡支付	中青年	7.83%
	老年	74.12%

使用郑州市公交乘客特性调查结果，可知老年（>60 岁）约占比 4.1%，中青年（>10 岁）约占比 93.3%[3]，其余年龄小于 10 岁人群占比较少，此处不予考虑。统计结果如下图所示。修正后，老年占比 4.2%青年占比 95.8%。

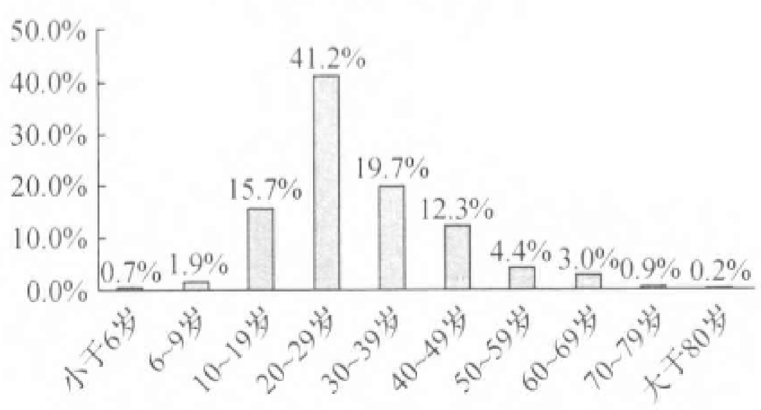


图 7-8 郑州公交调查年龄分布[3]

由此，计算全覆盖条件下，选择移动支付的概率为：

$$P_1=\frac{P_{1y}P_y+P_{1o}P_o}{P_{1y}P_y+P_{1o}P_o+P_{2y}P_y+P_{2o}P_o}=89.50\% \tag{7-4-10}$$

$$P_2=\frac{P_{2y}P_y+P_{2o}P_o}{P_{1y}P_y+P_{1o}P_o+P_{2y}P_y+P_{2o}P_o}=10.50\% \tag{7-4-11}$$

其中下标$\{1,2\}$代表决策选项（移动支付、公交卡支付），$\{y,o\}$代表人群（中青年、老年）。

7.5 求解结果

综合集计模型（创新扩散模型）与非集计模型（离散选择模型）概率结果，计算平均值。可得：全覆盖条件下，选择移动支付的概率约为 86.75%，选择公交卡支付的概率约为 13.25%。

将该选择概率带入问题二模型中，经过计算可得，实现全覆盖后，第三方平台每月约盈利：724.179 万元。

八、问题四的方案

8.1 第三方移动支付平台的商业可行性报告

随着工业水平和信息技术的不断进步，智能移动终端设备即智能手机已经走进了千家万户，而依托于其的第三方移动支付平台也拥有了更多的商业机会。通过与实体企业

进行合作，将第三方移动支付平台与已有的实体磁卡，如：银行卡、乘车卡、会员卡等进行功能的结合，并在使用的过程中进行手续费、服务费、广告费以及沉淀资金利息的收取，以达到盈利。

在本问题中，第三方支付平台通过与公共交通系统进行合作，将乘车卡的充值、使用等功能与平台结合。在具体的商业进程中，前期的宣传、基础设施的建设以及员工的劳务费都随着影响范围的扩大在不断增长，但是，用户数量的增加也扩大了平台的盈利能力，尤其是服务费和沉淀资金的利息的增长曲线远高于支出曲线。本文中有盈利模型：$W_0 = I_0 - O_0$，其中人数n增长，$f'(I_0) > f'(O_0) \to W_n - O_n > o \to W_n - O_n > W_{n-1} - O_{n-1}$，即盈利$W$在不断增长。本次问题中可知当平台占有率从$50\%$增至$86.75\%$时，利润从 190 万增至 720 万，具有极高的利润。

第三方支付平台的可行性建议：

①第三方支付平台可以扩大融资、增加平台的商业影响力和规模。

②第三方支付平台可寻求与更多实体企业的多方面合作。

③第三方支付平台可适当降低手续费，吸引更多客户。

④第三方支付平台可增加多种服务方向和便携式移动支付途径。

九、模型的评价与推广

问题一分析了附件给出的数据，用图表的形式展示出支付特征之间的关系，能够客观、可视化的衡量每个因素在数据表单中的作用，并且通过这个方式分析出该城市乘车人出行支付特征，并且为之后的问题做一个基础的数据支持。

问题二建立了第三方移动支付平台的数学模型，该模型具有可靠性并且客观、全面的结合了影响收入和支出的因素，并且将盈利模型细分为各个模块，再定量的解析得出盈利结果，提高了函数准确性和精度。

问题三建立创新扩散模型，将四分之一应用的数据推广到全部应用时的数据。并且利用解决供需关系的集计和非集计模型解释了使用这个模型的先决条件，科学完整的建立了一个新的模型并且结合了第二问的盈利模型。寻找客流量、第三方支付平台支付覆盖率、第三方支付企业盈利之间的关系，带入数据对模型进行验算。本问所用模型具有创新性和高效性，利用专业知识完成建模过程。

本文所建立的多种模型具有创新性，实际实用性强，充分利用了附件各数据及指标因素，从它们之间的性质与数量方面都做出了分析与评价，为不同的项目及操作方式提供了良好的定价方案，具有较广的适用范围。

十、参考文献

[1] Everett M.Rogers. Diffusion of Innovations. New York：The Free Press, 1983

[2]段鹏. 离散选择模型理论与应用研究[D].南开大学,2010.

[3]冯建栋,王昊.郑州市公交乘客特性及出行意愿调查分析[J].交通科技与经济,2015,17(04):64-70.

附 录

附录 1

文件：data mining1
环境：jupyter notebook
编程语言：Python
使用模块：numpy、pandas、matplotlib、seaborn、datetime

```
# coding: utf-8
# In[1]:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
get_ipython().magic('matplotlib inline')
# In[2]:
data_0207 = pd.read_excel('20170207.xlsx')
data_0208 = pd.read_excel('20170208.xlsx')
data_0209 = pd.read_excel('20170209.xlsx')
data_0210 = pd.read_excel('20170210.xlsx')
data_0211 = pd.read_excel('20170211.xlsx')
data_0212 = pd.read_excel('20170212.xlsx')
data_0213 = pd.read_excel('20170213.xlsx')
data_0511 = pd.read_excel('20170511.xlsx')
data_0512 = pd.read_excel('20170512.xlsx')
data_0513 = pd.read_excel('20170513.xlsx')
data_0514 = pd.read_excel('20170514.xlsx')
data_0515 = pd.read_excel('20170515.xlsx')
data_0516 = pd.read_excel('20170516.xlsx')
data_0517 = pd.read_excel('20170517.xlsx')
data_0810 = pd.read_excel('20170810.xlsx')
data_0811 = pd.read_excel('20170811.xlsx')
data_0812 = pd.read_excel('20170812.xlsx')
data_0813 = pd.read_excel('20170813.xlsx')
data_0814 = pd.read_excel('20170814.xlsx')
data_0815 = pd.read_excel('20170815.xlsx')
data_0816 = pd.read_excel('20170816.xlsx')
data_1109 = pd.read_excel('20171109.xlsx')
data_1110 = pd.read_excel('20171110.xlsx')
data_1111 = pd.read_excel('20171111.xlsx')
data_1112 = pd.read_excel('20171112.xlsx')
data_1113 = pd.read_excel('20171113.xlsx')
data_1114 = pd.read_excel('20171114.xlsx')
data_1115 = pd.read_excel('20171115.xlsx')
# In[17]:
data_0207 = data_0207.iloc[:,:4]
data_0208 = data_0208.iloc[:,:4]
data_0209 = data_0209.iloc[:,:4]
data_0210 = data_0210.iloc[:,:4]
data_0211 = data_0211.iloc[:,:4]
data_0212 = data_0212.iloc[:,:4]
data_0213 = data_0213.iloc[:,:4]
data_0207.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0208.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0209.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
```

```
data_0210.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0211.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0212.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0213.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0213.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0511.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0512.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0513.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0514.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0515.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0516.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0517.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0810.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0811.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0812.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0813.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0814.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0815.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0816.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1109.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1110.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1111.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1112.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1113.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1114.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_1115.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_ = pd.concat([data_0207,data_0208,data_0209,data_0210,data_0211,data_0212,data_0213,
data_0511,data_0512,data_0513,data_0514,data_0515,data_0516,data_0517,
data_0810,data_0811,data_0812,data_0813,data_0814,data_0815,data_0816,
data_1109,data_1110,data_1111,data_1112,data_1113,data_1114,data_1115],axis=0)
print(data_.shape)
# In[18]:
data_.head(10)
# In[19]:
data_['UPTIME'] = data_['UPTIME'].replace({'0001-1-1':'1900-01-01 00:00:00'})
# In[21]:
data_ = data_[(data_['UPTIME']!='1900-01-01 00:00:00')]
# In[22]:

data_.head()
# In[23]:
data_['Date'] = pd.to_datetime(data_['UPTIME'])
data_['Month'] = data_['Date'].dt.month
data_['Day'] = data_['Date'].dt.day
data_['Weekday'] = data_['Date'].dt.weekday
data_['hour'] = data_['Date'].dt.hour
# In[24]:
data_['PAYTYPE']= data_['PAYTYPE'].replace({'Null': np.nan})
# In[25]:
data_ = data_[(data_['PAYTYPE']<=1)]
# In[26]:
data_.head(10)
# In[27]:
a=data_.groupby(['Day','hour'])['PAYTYPE'].value_counts()
plt.plot(np.arange(24),a[:,0],label='onlinepay')
plt.plot(np.arange(24),a[:,1],label='card')
plt.legend()
```

```
plt.xlabel('Hour')
plt.xticks(np.arange(24))
plt.savefig("Hour.png");
# In[57]:
b=data_.groupby(['Weekday'])['PAYTYPE'].value_counts()
b[:,0].plot(label='onlinepay')
b[:,1].plot(label='card')
plt.legend()
plt.xlabel('Weekday')
plt.xticks([0,1,2,3,4,5,6],[r'$Mon$',r'$Tue$',r'$Wed$',r'$Thur$',r'$Fri$',r'$Sat$',r'$Sun$'])
plt.ylim(1600000,1700000)
plt.savefig("Weekday.png");
# In[31]:
c = data_.groupby(['Month'])['PAYTYPE'].value_counts()
# In[39]:
portion=c[:,0]/(c[:,0]+c[:,1])
# In[41]:
portion
# In[56]:
y1=np.array([0.497243,0.496235, 0.498394,0.500093])
y2=np.array([1.0,1.0,1.0,1.0])-y
ind = np.arange(4)                    # the x locations for the groups
width = 0.3
plt.bar(ind,y1,width,label = 'onlinepay')
plt.bar(ind+width,y2,width,label = 'card') # ind+width adjusts the left start location of the bar.
plt.xticks([0,1,2,3],[r'$Feb$',r'$May$',r'$Aug$',r'$Nov$'])
plt.xlabel('Month')
plt.ylabel('onlinepay portation')
plt.ylim(0.3,0.55)
plt.legend()
plt.savefig("Month.png");
# In[46]:
# In[61]:
c[:,0].plot(label='onlinepay')
c[:,1].plot(label='card')
plt.legend()
plt.xlabel('Month')
plt.xticks([2,5,8,11],[r'$Feb$',r'$May$',r'$Aug$',r'$Nov$'])
plt.ylim(2200000,3500000)
plt.savefig("month2.png");
# In[65]:
data_.head()
# In[66]:
data_ = data_[(data_['LASTTIME']!='0001-1-1')]
# In[84]:
data_['Lastdate'] = pd.to_datetime(data_['LASTTIME'],errors='coerce')
# In[93]:
data_.head()
# In[94]:
data_['delta']=data_['Date'] - data_['Lastdate']
# In[101]:
data_.head()
# In[138]:
guest = data_.groupby('ID').agg({'Month': len, 'PAYTYPE': lambda x:
len(x[x==0])}).rename(columns={'Month':'Count','PAYTYPE':'card'})
# In[144]:
guest.columns = ['Count','Onlinepay']
```

```
# In[153]:
guest['Count'].describe()
# In[159]:
guest=guest[guest['Count']<200]
# In[160]:
guest['Count'].describe()
# In[167]:
sns.distplot(guest['Count'],bins=50)
plt.savefig("hist.png");
# In[168]:
guest.shape[0]
# In[171]:
sns.distplot(guest['Onlinepay'],bins=50)
plt.savefig("Onlinepayhist.png");
# In[172]:
guest['card']=guest['Count']-guest['Onlinepay']
# In[173]:
sns.distplot(guest['card'],bins=50)
plt.savefig("cardhist.png");
# In[174]:
guest['onlineportion']=guest['Onlinepay']/guest['Count']
# In[177]:
sns.distplot(guest['onlineportion'])
plt.xlabel('online ratio')
plt.ylabel('count')
plt.savefig("onlineportion.png");
# In[180]:
plt.hist(guest['onlineportion'],bins=40,facecolor='g',alpha=0.5)
plt.xlabel('online ratio')
plt.ylabel('count')
plt.savefig("onlineportion1.png");
# In[188]:
guest1=guest[guest['Count']>=50]
plt.hist(guest1['onlineportion'],bins=40,facecolor='darkblue',alpha=0.5)
plt.xlabel('online ratio')
plt.ylabel('count')
plt.savefig("onlinecountbig.png");
# In[190]:
data_['Date_ex']=data_['Date'].dt.date
# In[204]:
y=data_.groupby(['Date_ex'])['PAYTYPE'].value_counts()
ii=y[:,0]
jj=y[:,1]
ii[ii>100].plot(label='onlinepay')
jj[jj>100].plot(label='card')
plt.legend()
plt.xlabel('Date')
plt.xticks(rotation=90)
plt.savefig("Date.png");
# In[208]:
data_['deltahour']=data_['delta'].dt.seconds//3600
# In[211]:
data_['deltaday']=data_['delta'].dt.days
# In[228]:
data_['deltahour'].describe()
```

```
# In[232]:
data_['deltahour'].hist(bins=23,facecolor='b',alpha=0.3)
# In[236]:
data_.head()
# In[238]:
z=data_.groupby(['deltahour'])['PAYTYPE'].value_counts()
# In[235]:
pp=z[:,0]
qq=z[:,1]
plt.legend()
plt.xlabel('delta time')
plt.xticks(rotation=90)
plt.savefig("deltatime.png");
```

附录 2

文件 2：data mining2
环境：jupyter notebook
编程语言：Python
使用模块：numpy、pandas、matplotlib、seaborn、datetime

```
# coding: utf-8
# In[1]:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
get_ipython().magic('matplotlib inline')
# In[2]:
data_0207 = pd.read_excel('20170207.xlsx')
data_0208 = pd.read_excel('20170208.xlsx')
data_0209 = pd.read_excel('20170209.xlsx')
data_0210 = pd.read_excel('20170210.xlsx')
data_0211 = pd.read_excel('20170211.xlsx')
data_0212 = pd.read_excel('20170212.xlsx')
data_0213 = pd.read_excel('20170213.xlsx')
# In[4]:
data_0207 = data_0207.iloc[:,:4]
data_0208 = data_0208.iloc[:,:4]
data_0209 = data_0209.iloc[:,:4]
data_0210 = data_0210.iloc[:,:4]
data_0211 = data_0211.iloc[:,:4]
data_0212 = data_0212.iloc[:,:4]
data_0213 = data_0213.iloc[:,:4]
data_0207.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0208.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0209.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0210.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0211.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0212.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_0213.columns = ['ID', 'LASTTIME', 'UPTIME', 'PAYTYPE']
data_ = pd.concat([data_0207,data_0208,data_0209,data_0210,data_0211,data_0212,data_0213],axis=0)
# In[5]:
data_ = data_[(data_['UPTIME']!='0001-1-1')]
data_['Date'] = pd.to_datetime(data_['UPTIME'])
data_['Month'] = data_['Date'].dt.month
data_['Day'] = data_['Date'].dt.day
```

```
data_['Weekday'] = data_['Date'].dt.weekday
data_['hour'] = data_['Date'].dt.hour
# In[6]:
data_['PAYTYPE']= data_['PAYTYPE'].replace({'Null': np.nan})
data_ = data_[(data_['PAYTYPE']<=1)]
# In[7]:
data_.head()
# In[16]:
imdata=data_.groupby(['Day','hour'])['PAYTYPE'].value_counts()
# In[17]:
imdata2 = np.zeros((24,7))
# In[21]:
imdata
# In[26]:
for i,k in enumerate(range(7,14)):
	for j in range(24):
		imdata2[j,i]=imdata[k,j,1]
# In[35]:
plt.figure(figsize=(20,14))
plt.imshow(imdata2,cmap='Purples')
plt.xticks([0,1,2,3,4,5,6],[r'$2-7$',r'$2-8$',r'$2-9$',r'$2-10$',r'$2-11$',r'$2-12$',r'$2-13$'],rotation=45)
plt.yticks(np.arange(0,24,1))
plt.title('Use card to pay in Feb')
plt.xlabel('Date')
plt.ylabel('Hour')
plt.colorbar()
plt.savefig('Feb.png');
# In[38]:
imdata3 = np.zeros((24,7))
for i,k in enumerate(range(7,14)):
	for j in range(24):
		if (k==8)&(j==3):
			imdata3[j,i]=0
		elif (k==10)&(j==3):
			imdata3[j,i]=0
		else:
			imdata3[j,i]=imdata[k,j,0]
plt.figure(figsize=(20,14))
plt.imshow(imdata3,cmap='Purples')
plt.xticks([0,1,2,3,4,5,6],[r'$2-7$',r'$2-8$',r'$2-9$',r'$2-10$',r'$2-11$',r'$2-12$',r'$2-13$'],rotation=45)
plt.yticks(np.arange(0,24,1))
plt.title('Use OnlineAPP to pay in Feb')
plt.xlabel('Date')
plt.ylabel('Hour')
plt.colorbar()
plt.savefig('Febonlie.png');
```