



同濟大學
TONGJI UNIVERSITY

2020-2021 学年运输经济学大作业
共享单车出行分析与预测

姓名: 徐晨龙

学号: 1854084

专业方向: 交通信息

专业学院: 交通运输与工程学院

引言：无桩共享单车逐渐成为城市交通中不可获取的一部分，本文通过以获取的新加坡共享单车数据和城市交通的其他基本参数，分析共享单车投放量和出行量之间的关系，并尝试利用边际效应来解释贡献共享单车投放量和使用量之间的关系；同时通过建模来分析其他参数，如交通基础设施、城市建成环境和天气等数据对共享单车投放量的影响，并在上述工作的基础上，对共享单车骑行需求进行预测，讨论共享单车这一运营模式的可行性。

一、数据情况简介与描述性分析

1. 共享单车需求影响因素分析

在能源危机出现、全球变暖等问题下，社会需求的驱动下，以自行车和步行为代表的慢性交通逐渐回归，伴随着共享经济的进步，共享单车逐渐在中国的北上广深等城市脱颖而出，也逐渐进驻新加坡等国家。影响城市居民对共享单车的选择因素主要有以下两种^[1]

（1）微观因素，共享单车的使用必然和出行者的基础属性和出行特征有关，基础属性常为性别、年龄、受教育程度和收入水平；出行特征主要为出行目的地和出行费用有关。

（2）宏观因素，共享单车系统作为开放性系统，宏观因素也对共享单车的影响作用分析很重要，主要有共享单车投放量，以及城市基础设施的基本属性，本文数据中主要考虑：与地铁站之间的距离、ID 点的香农熵、自行车车道产长度、公有住宅建筑/私有住宅建筑/商业建筑容积率

2. 数据来源分析

截至 2017 年，共享单车在全球 1328 城市中，有桩共享单车总共约 134 万辆，但在无桩共享单车发展的情况下，今上海，无桩共享单车投放量约为 178 万辆，中国共投放超过 2000 万辆，之后 ofo、mobike 也逐渐进入新加坡市场，本文主要通过已获取的新加坡共享单车 GPS 数据和公共基础交通数据进行分析

截至 2016 年，新加坡是一个城市国家，总面积 $721.5km^2$ ，人口 560 万，是世界上人口稠密和城市化程度最高的地区只有，具有集中的公共交通系统^[2]，空气好，污染少，同时新加坡政府一致致力于城市建设和发展进程中可持续发展理念，并在 2013 年推出 Nation Cycling Plan 计划，但在另一方面由于自然环境和共享

单车相关管理并不完善。新加坡只有旱雨季之分，旱季炎热潮湿、雨季多雨，并不适合骑行，参照图 1，可以看出本文所选数据 8 月份为高温潮湿的旱季，相关交通法规并不完善，对骑行者的年龄要求、基本条件、脚踏车基本配置和路权的优先次序无明确规定，容易造成很大的交通安全隐患。

High & Low Weather Summary for August 2017

Temperature		Humidity	Pressure
High	33 °C (1 Aug, 15:00)	100% (17 Aug, 17:23)	1013 mbar (17 Aug, 17:23)
Low	24 °C (17 Aug, 17:00)	52% (30 Aug, 14:00)	1005 mbar (26 Aug, 16:00)
Average	29 °C	76%	1009 mbar

* Reported 1 Aug 05:00 — 31 Aug 23:00, Singapore. Weather by CustomWeather, © 2020

图 1-1: weather summary for August 2017 in Singapore

3. 零值筛选

由于所给数据已经经过 GPS 数据路径、自行车速度和异常数据处理，因此仅进行零值数据筛选即可，筛选思路为：如果起点 m 到终点 n 有骑行数据，说明两点之间是可以骑行的，即可能存在 OD 对 (m,n) 或者 (n,m) 。因此筛选过程是根据两点之间是否存在骑行数据，并进行 OD 交换得到所有可能出现的 OD 对，再对原始数据中 OD 删去不可能出现的 OD 对，得到对应数据：

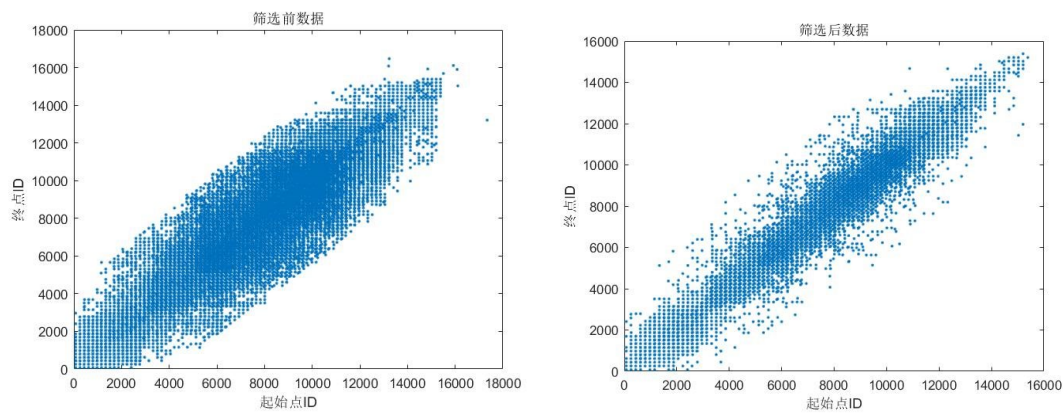


图 1-2: 零值筛选前后数据

由图 1-2 可以看出，从原始数据中排除 GPS 漂移短途旅行、共享单车搬运长度旅行以及 OD 对之间的合理距离之后，仍有可能出现冗余数据，之后为了判断两 OD 对可以通过两 OD 对是否可行来判断，但本次数据采取的是仅通过一个月的数据来判断是否为可达 OD 对，结果可能会有偏差。

4. 数据总览

新加坡八月份工作日早高峰期间 OD 对的平均骑行总数共有 219386 车次，根据产生的 OD 对以及路网距离可以得到骑行的距离分布：

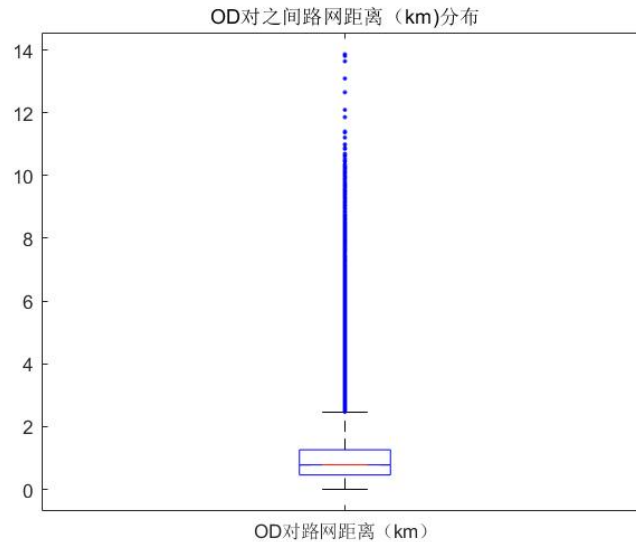


图 1-3：OD 对路网距离分布

$Q1=0.46\text{Km}$ ， $\text{median}=0.7826\text{Km}$ ， $Q3=1.26\text{km}$ ，

下四位数为 0.46Km ，上四分位数位 1.26Km ，中位數位 0.78Km ，其中上邻为 2.46Km ，下邻为 0Km 。由图可以看出 OD 对之间的距离分布常见值在 $(0, 2.46)\text{km}$ 之内，同时根据箱型可以看出自行车出行距离分布较为集中在短途。

同时由各点的共享单车投放量和共享单车出行数量可以得到相应的分布图

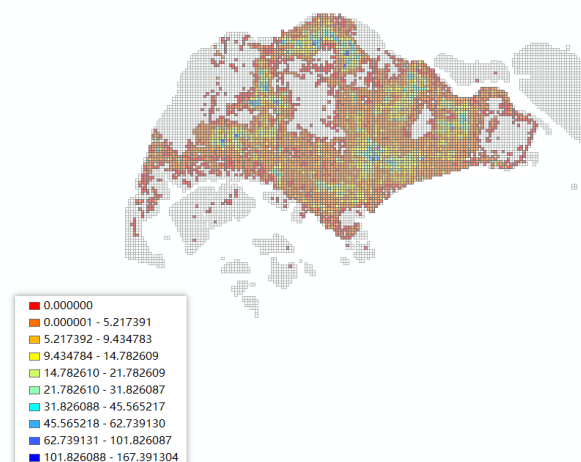


图 1-4：共享单车投放量空间分布

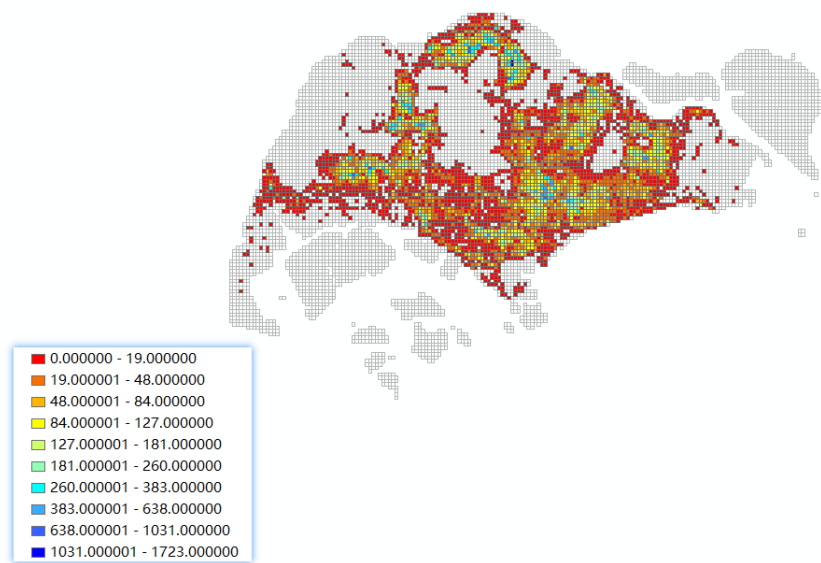


图 1-5：共享单车空间生成量

图中灰色表示没有数据，由两图可以初步看出共享单车投放量和 OD 对之间的骑行是成正相关的关系，即共享单车投放量越多，潜在的客户就越多。

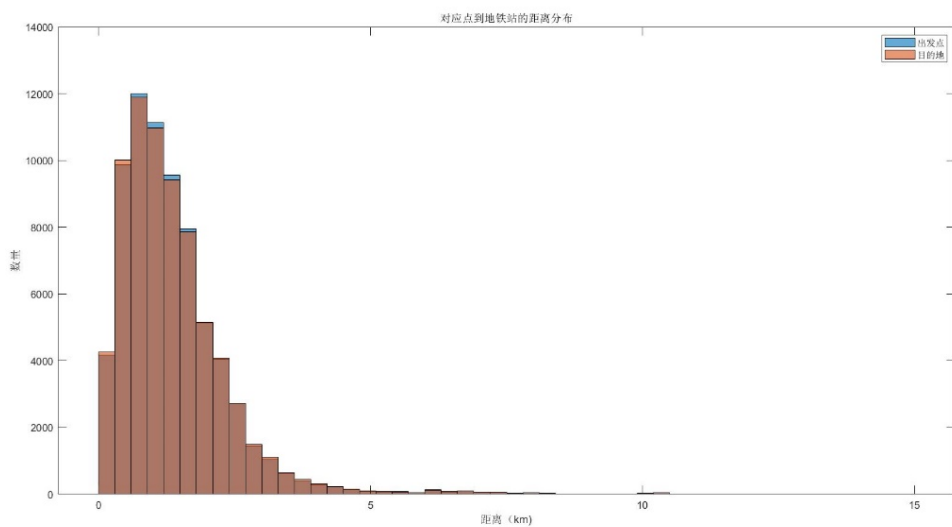


图 1-6 对应点到地铁站的距离分布

由图可以看出两者的距离分布都较为接近，与共享单车骑行距离主要发生在 1~2km 之外。

二、共享单车 O 点投放量与骑行量之间的关系

1. 单个 OD 对共享单车投放量和骑行量之间的建模

假设共享单车初始量为 p_0 ，共享单车投放量为 p ，共享单车的骑行量为 q ，出发点内共有 N_0 个人群，每个人选择共享单车的概率为 $p_i(0 < i \leq N_0)$ ，那么则有

$$q = \sum (p + p_0) * p_i$$

首先，由于物理空间的限制、交通的监管以及经济等因素，共享单车的投放量不能超过某一个定值即

$$p < p_{max}$$

同时显然，人群选择共享单车会受到各种因素的影响，常见的有如下：

影响因素	影响方面
年龄因素	不同年龄的个体具有不同的运动能力，影响是否选择共享单车
经济因素	新加坡共享单车租金较贵，影响对共享单车出行的选择
目的地选择	如果目的地较远，可能不会选择共享单车出行
道路条件	某种条件难满足自行车行驶的需求
时间因素	因为共享单车行驶速度限制，影响对其选择
...	...

由上可以看出选择概率受到大量独立的、均匀微小的随机因素的叠加影响，假设因其他因素造成骑行的记为 1，造成放弃骑行的即为-1，在人群足够多的情况下，那么根据中心极限定理，我们有理由相信这些因素满足正态分布，并且其均值为 0，可以得出

$$\Delta p_{i_x} = a * \exp(-x^2/b)$$

假同时由于人群的无记忆性，即每个人选择共享单车出行的概率并不会因为共享单车的数量发生变化，即每个人选择共享单车的概率：

$$p_x(p > p_1 + \Delta p) - p_x(p > p_1) = p(p > \Delta p)$$

因此可以推测人群选择共享单车与共享单车投放量之间的关系满足负指数分布：

$$p_x = a * \exp(-(p - p_{max})/c)$$

由上述推导过程可以得出

$$q = \sum (p + p_0) * (p_i + \Delta p_{ix}) \approx p * a * N0 * \exp(-(p - p_{max})/c))$$

在假设初始没有共享单车投放的情况，可以将投放量和骑行量之间的关系化简为上式子.

2. 数据拟合

将经过 1-3 中零值处理之后的数据进行窗口为 30 的移动平滑处理，尽量去除不可靠的异常值，便于观察数据的大体趋势：

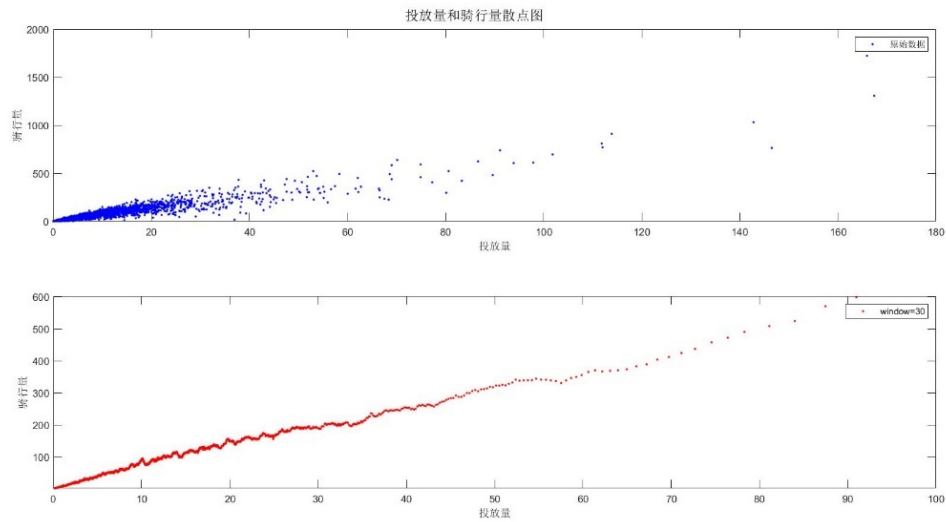


图 2-1：投放量和骑行量散点图（原始数据和平滑之后）

之后利用 matlab 的 curve fitting 工具箱进行函数拟合，可以得到

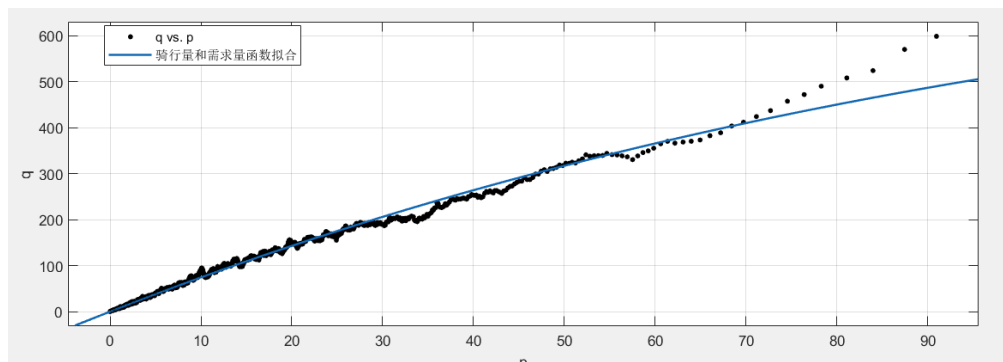


图 2-2：平滑后的骑行量和需求量的拟合

$$q(p) = a * p * \exp(-((p - b)/c))$$

在 95%的置信区间下，各参数的取值区间：

$$a = 5.84 \quad b = 51.34 \quad c = 129.8$$

相关参数:

$$R - square = 0.9854$$

$$RMSE = 8.015$$

可以看出拟合结果在函数左半段较为理想，随着 p 接近 p_{max} ，函数预测会越来越不准，可能存在的原因是随着共享单车数量的逐步提升，市场上可供选择的其它共享单车品牌会逐渐减小，导致选择该品牌共享单车的数量会越来越多，逐渐形成垄断市场。

3. 单个 O 点投放量和骑行生成量的边际函数

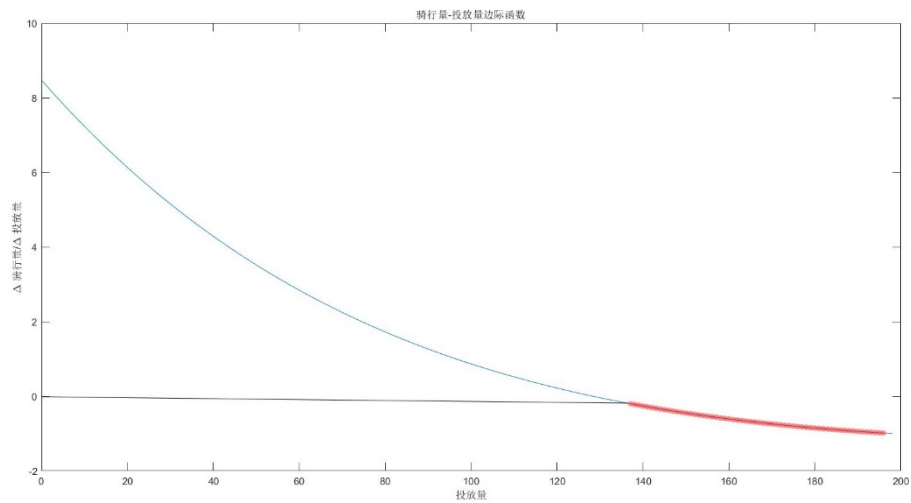


图 2-3：骑行量和投放量的边际函数

可以看出随着投放量的增大，每单位投放量引起骑行量变化都会逐渐减小，刚开始投入的共享单车的边际效益最大，当单个 O 点的共享单车投放量达到 128 之后（图 2-3 的红色部分），每单位投放量会引起骑行量变化的减少，这从实际上不可能的，这一部分是由于没有考虑除了共享单车需求量以外的因素导致的。即 p_{ix} 的影响造成的。

4. 整个地区的 O 点投放量和共享单车使用量模型

由于单个 OD 对的共享单车投放量和使用量之间的关系，由于共享单车投放不均匀的特性，不能代表整个新加坡取 6-9 月的共享单车投放量和使用量的数量，因此从所得到的数据中提取出 6~9 月早晚高峰的（共享单车 O 点总投放量，共享单车使用量）的数据：（22705.19，124003）（38802.64，256389）

（28531.52，173315）（42471.28，288991）（30357.49，219306）（40874.25，

323649) (25780.33, 148201) (37037.42, 236564)

同时发现可以生成 OD 对的 O 点大约有 4000 个左右，可见平均每个单独点的共享单车投放量不大于 10，此时边际效益大约在【10，8】之间，出于较稳定的状态，因此可以假设一个月份中整个地区的共享单车投放量和共享单车骑行生成量成线性关系：

$$trip_{total} = a * fsize_{total} + b$$

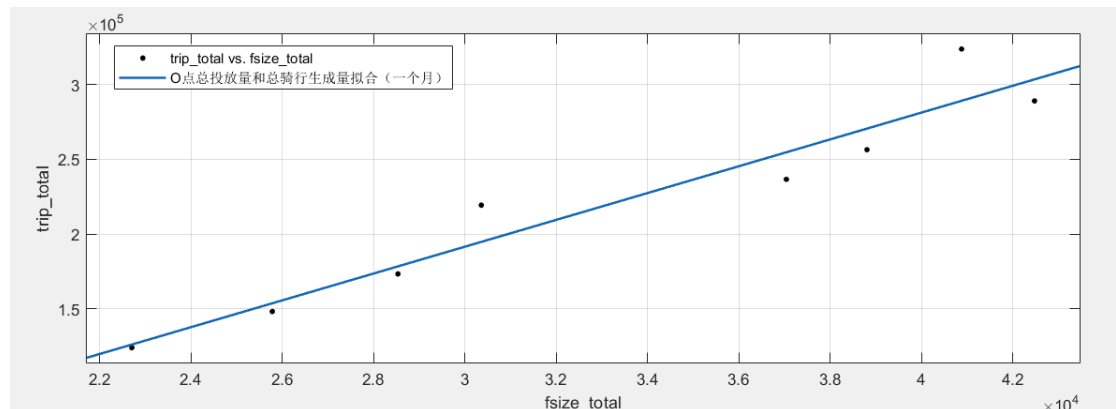


图 2-4:一个月内 O 点投放量和共享单车骑行总量关系拟合

拟合结果：

$$a = 8.936 \quad b = -77550$$

$$R\text{-square} = 0.9229$$

其结果需求与初步估计结果相同，因此可认为符合条件，因此对于一个月新加坡地区 O 点的总共享单车投放量 $fsize_{total}$ 和单次共享单车骑行生成总量 $trip_{total}$ 的函数为：

$$trip_{total} = 8.836 * fsize_{total} - 77550$$

三、共享单车出行影响因素分析

1. 定性分析交通基础设施、城市建成环境等因素等影响

所给数据中交通基础设施主要包括到附近地铁站的距离和城市自行车车道长度。起点和终点附近地铁站距离影响出行的可能原因在于，出行者刚刚从地铁出站利用共享单车接驳，或者骑共享单车到达出发点，这两者的概率应该是随机的；城市自行车道长度越完善，更加有利于人们出行选择自行车，进而会促进人们选择共享单车，提高其需求。

城市建成环境的影响参考建筑环境和交通^[3]的影响，通常在自行车相关研

究中会评估土地使用和街道配置，其中利用土地面积占比来表示土地利用密度，分别给出了公共住宅容积率、私有住宅容积率和商业用地容积率，另一方面利用POI 计算香农熵来衡量土地利用的混合程度，可预见的是土地利用率高，城市经济越发达，人群出行需求也就越多，共享单车需求也会增大，但同时也会使得自行车出行不便利。

$$H = - \sum_i p_i \log_n p_i$$

2. 数学模型的建立

通过绘图可以得到

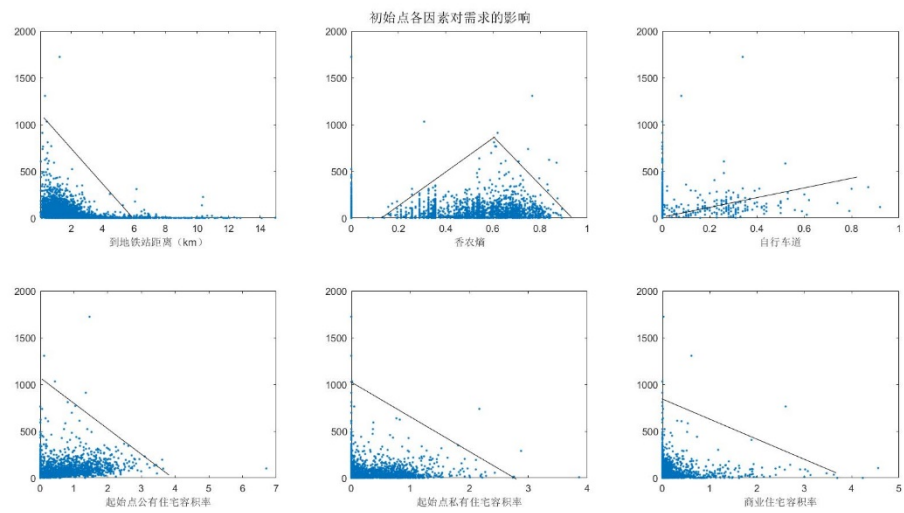


图 3-1：初始点各因素对需求的影响

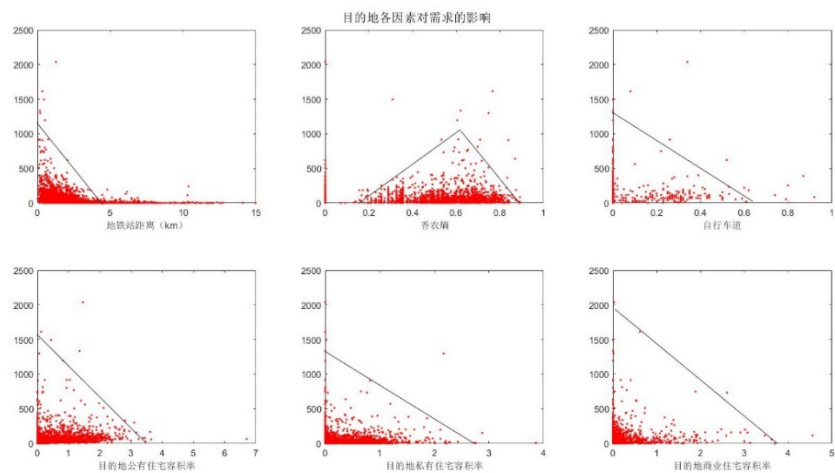


图 3-2：目的地各因素对需求的影响

由于影响因素对需求的影响是最外侧包络部分，因为可能有的点并没有充分的将需求释放出来，通过绘制出现有各因素外侧包络可以观察出因素对需求的影响。

首先分析初始点：从图中能较为明显的看出，离地铁站较近得地方，共享单车生成量较高，可以理解为早高峰时期，从地铁站出站后骑行共享单车得占据大多数，香农熵位于中间区段对需求的影响居多，香农熵过小或者过大都会影响共享单车的出行需求。起始点附近自行车道、公有住宅、私有住宅、商业化的容积率对共享单车的变化影响并不明显，其趋势是随着容积率的上升，会导致共享单车需求减少，但并不是特别明显；具有自行车道德起始点，随着车道增加，共享单车趋势在逐渐上升。

目的地上述因素影响和初始点的影响大致相同。结合 2-1 中的建模思路，得到需求量 Q 与投放量 P 、距离地铁站距离 L_1 、起始点香农熵 H_1 和终点的香农熵 H_2 、附近自行车道 L_2 以及与不同容积率之间的不计算的概率 $f(x_1, x_2, x_3)$

$$Q = N_0 * (p + p_0) * [a_1 * \exp(-(x - b_1)/c_1) + a_2 * L_1 + a_3 * \exp(-(H_1 + H_2 - b_3)^2/c_2) + a_4 * L_4 + c + f(x_1, x_2, x_3)]$$

3. 各种因素弹性分析

运输需求弹性常用来分析需求随其影响因素变化而变化的反应程度，即是运输需求量变化的百分率与影响运输需求的因素变化百分率的比值。

$$E_x = (dQ/dx) * (x/Q)$$

投放量对需求的弧弹性，取 (0, 0), (90, 600)；

$$E_p = 1$$

初始点地铁站距离对需求的弧弹性，取 (0.17, 911), (7.83, 12)：

$$E_{L1} = -1.017$$

香农熵对需求的弧弹性：(0.1861, 30), (0.6198, 911), (0.8918, 4)

$$E_H^1 = 1.74 \quad E_H^2 = -5.5$$

自行车道对需求的弧弹性：(0.00839, 41), (0.8, 313)

$$E_{L2} = 0.79$$

容积率对需求的弧弹性：

$$E_{x1} = -0.6 \quad E_{x2} = -0.94 \quad E_{x3} = -0.99$$

需求弹性根据绝对值得影响，可以得到 $E_x = 0$ ，表示完全无弹性， $E_x < 1$ 表示缺乏弹性， $E_x = 1$ 表示单位弹性， $E_x = \infty$ 表示完全有弹性。则根据计算结果，每增加1%投放量会对应增加1%的需求量；离地铁站距离每增加1%会减少1.017%的需求量，香农熵每在上升段每增加1%引起需求量增加1.74%，下降端每下降1%引起需求量增加5.5%，容积率每减少1%，对应增加需求量（0.6%，0.94%，0.99%），可以看出香农熵对于自行车需求量增加的弹性最大，尤其是将土地多样性维持在一个较为稳定的水平。

四、共享单车的需求预测模型

交通需求是社会经济活动在人与货物空间位移方面所提出的有支付能力的需要，需求预测的常见基本原理有：惯性原理、类推原理和相关原理。在共享单车需求分析中可以看出有较多稳定不变的因素，其变化具有随季节变化的周期性和随着基础设施和投放量的缓慢稳定变化的趋势，因此在对6-9月中的运量的基础上，通过减去平均值后拟合，得到在投放量、基础设施等不发生改变情况下，早高峰共享单车需求预测函数：

原始数据：(6, 124003)、(7, 173315)、(8, 219386)、(9, 148201)

$$median = 160758$$

$$M(1,2,3,4)=[124003 \ 173315 \ 219386 \ 148201]-median*ones(1,4)$$

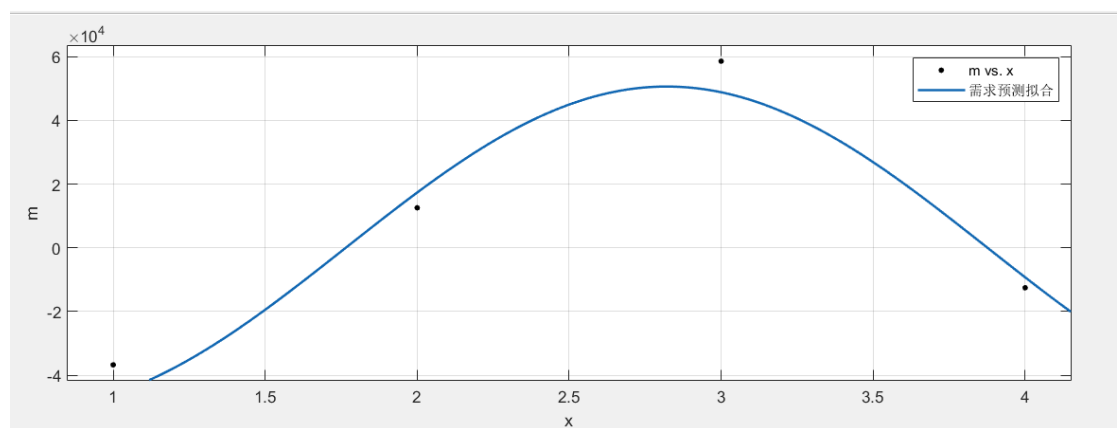


图 4-1：需求差值拟合函数

拟合结果 $R - square = 0.957$ ，较为理想。

$$Q(t) = median + m(x) = 160758 + 50650 * \sin(1.488 * (t - 6) + 3.657)$$

注：t:为月份，每跨一年加 12

根据 2-4 中月份共享单车总投放量和总骑行量生成的关系,如果共享单车增加 ΔP

$$Q(t, \Delta P) = 160758 + 8969 * \Delta P + 50650 * \sin(1.488 * (t - 6) + 3.657)$$

即共享单车月份需求总量除了和时间有关，也和共享单车供给之间有关，其随月份出现周期性的波动，随共享单车的攻击者增加而增加。

五、结果分析和讨论

通过分析以月份为跨度的共享单车骑行量和单 O 点、总 O 点共享单车投放量，共享单车需求量和投放量、交通设施建设和城市土地空间的关系，可以得到需要提升共享单车需求量的主要措施有：提升投放量、增加地铁站密度、修建自行车道、维持香农熵、减小空间容积率。但在土地狭小的新加坡后两者方案并不可行，从图 5-1 的地铁网络可以看出新加坡地铁修建工作尚未结束，，有利于共享单车需求量的增加。

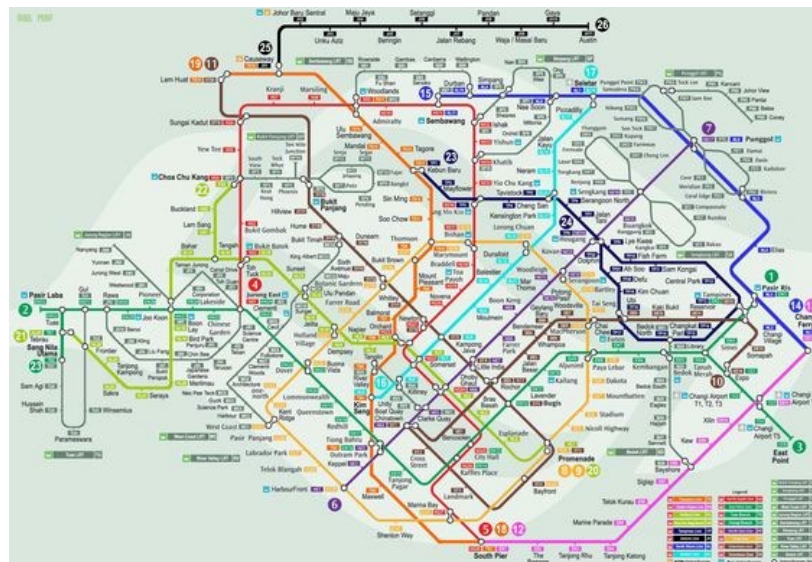


图 5-1：新加坡地铁网络

另一方面，可行 OD 对共享单车投放量大多集中于 40 一下，并没有达到所建模型中的边际效益小于 1 的点，因此继续增加早高峰对目标点的共享单车投放是可以增加出行需求量的。

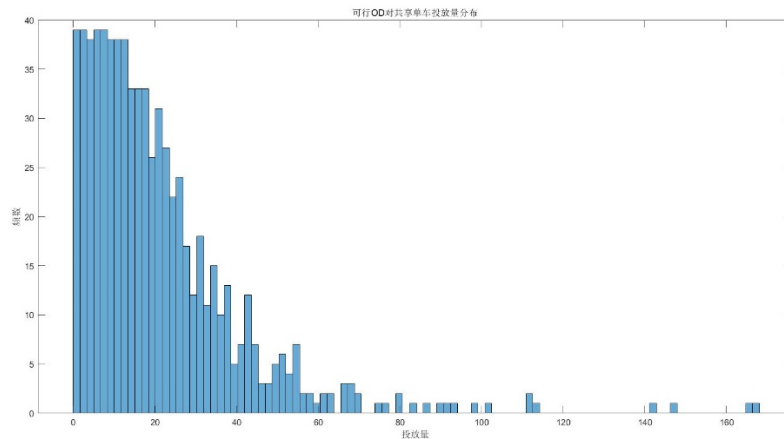


图 5-2：可行 OD 对共享单车投放量

六、总结

本文首先借助 2017 年 8 月份得新加坡共享单车数据，分析初步分析共享单车需求量影响得微观和宏观因素，利用判断 OD 对之间是否可达来对数据进行筛选，由于本文仅 1 个月得数据，OD 对得可达性可能并不完全正确，有待考证。之后从 OD 对路网距离分析、共享单车投放量热力图、共享单车空间生成量和距离地铁站距离来分析大体情况，发现共享单车出行集中在地铁站附近。之后通过理论推导得到共享单车骑行量和投放量之间的关系，并经过以移动平均之后通过拟合得到相应函数，所得结果中末端并不理想，可能原因时其他因素的影响未完全考虑。之后利用所得数据分析投放量、交通基础设施、城市建成环境对共享单车需求量进行分析并得到对应的弹性分析。结合所学知识讨论新加坡共享单车运营模式可行的原因。

参考文献:

- [1] 李玉. 城市共享单车选择行为影响因素研究[J]. 智能城市, 2020, 006(007):P.177-178.
- [2] Xu, Y., Chen, D., Zhang, X., Tu, W., Chen, Y., Shen, Y. and Ratti, C., 2019. Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. *Computers, Environment and Urban Systems*, 75, pp.184-203.
- [3]Ewing, R., & Cervero, R. (2010). Travel and the built environment: a meta-analysis. *Journal of the American planning association*, 76(3), 265-294.
- [4] 王丽. 共享单车发展中存在的问题与对策研究[D].山东师范大学,2019.