PSM匹配样本情况：

一共18个case，Treated : Control=3:15，ratio=1

最后从15个control组中选出3个

Sample sizes:

Control Treated

All 15 3

Matched 3 3

Unmatched 12 0

Discarded 0 0

匹配过程

1. 数据准备
   1. 选用了15个无公交车停靠的地点，3个有公交车停靠的地点，按照有无公交车为分类变量，匹配目的是在15个无公交车地点中选取三个作为对应的三个有公交车停靠的地点的代替。用PSM来减小主观因素。
   2. 按照IS\_BUS作为控制组（Control, WITH\_BUS=0）和实验组(Treated, WITH\_BUS=1)

Control组，WITH\_BUS=0的是根据匹配选出的三个location。

* 1. 选取的指标变量
     1. SEAT 公交站容量
     2. IS\_HARBOR ：是否是港湾式（1/0）
     3. Q\_AUTO：社会车辆流量 1hr
     4. Q\_NON\_AUTO： 非机动车流量 1hr
     5. STOP\_LOCATION: 公交站位置 近中远（-1/0/1）

1. 过程

在R语言下面载入“MatchIt” package， 使用其中的matchit方法，并采用“nearest”（, nearest neighbor matching）的匹配方法、采用logit对数据计算distance，进行匹配，其匹配主函数在R中的操纵脚本如下

matchit(WITH\_BUS~ SEAT+IS\_HARBOR+Q\_AUTO+Q\_NON\_AUTO+STOP\_LOCATION,data=mydata,method = "nearest",ratio=1,distance = "logit")

匹配结果 列表如下

The matching result (Treated sample=3, ratio=1):

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Matching result |  |  |  |  |  |  |  |  |
| location | SEAT | IS\_HARBOR | Q\_AUTO | Q\_NON\_AUTO | STOP\_LOCATION | IS\_BUS | distance | weights |
| location1 | 1 | 0 | 768 | 888 | -1 | 0 | 0 | 1.0000 |
| location3 | 1 | 0 | 840 | 1590 | -1 | 0 | 0 | 1.0000 |
| location7 | 3 | 0 | 1350 | 1368 | -1 | 0 | 0 | 1.0000 |
| location16 | 1 | 0 | 624 | 1032 | -1 | 1 | 0 | 1.0000 |
| location17 | 3 | 0 | 1398 | 1236 | -1 | 1 | 0 | 1.0000 |
| location18 | 1 | 0 | 1008 | 1230 | -1 | 1 | 0 | 1.0000 |

除了最近邻方法，也可以采用"exact" (exact matching), "full" (full matching), "optimal" (optimal matching), "subclass" (subclassification), and "genetic" (genetic matching)等匹配方法， 本文采取nearest方法， （weights可以不用放进去，跟匹配效果无关，由于所有的论文都没有提到weight，因此不是很明白 ，网上一种说法是说weight跟QQ图有关）

The propensity score model coefficients (ESTIMATE Std. Err.)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| The propensity score model |  |  |  |  |
|  | Coefficients(Estimate Std. Error) | | z | Pr(>|z|) |
| Intercept | -18.310 | 3765.000 | -0.005 | 0.996 |
| SEAT | -0.219 | 0.770 | -0.284 | 0.776 |
| IS\_HARBOR | NA | NA | NA | NA |
| Q\_AUTO | 2.566E-04 | 2.265E-03 | 0.113 | 0.910 |
| Q\_NON\_AUTO | -9.888E-05 | 2.147E-03 | -0.046 | 0.963 |
| STOP\_LOCATION | -1.716E+01 | 3.765E+03 | -0.005 | 0.996 |

（上表标明了所有变量对应的logit系数，由于用R处理PSM匹配时，只要数据和逻辑弄好了，是可以一步得出最后结果的，因此这里按照那篇博士论文列出的系数，在R中使用了“glm”包，把对应的系数单独用logit算了出来。横行对应的是各个变量对应的logit系数，具体logit系数含义分析由于在多篇论文中没找到解释，像那篇博士论文它那里边直接把这个系数放进去了，未进行详细解释，因此我这也不是太明白具体是什么含义T\_T，因此这里还得麻烦师兄您了。）

Check the covariates balance between group and after using nearest neighbors matching

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | sample | Mean | | SD Control | Mean Diff | eQQ Med | eQQ Mean | eQQ Max |
| Treated | Control |
| SEAT | Unmatched | 1.6667 | 1.8667 | 0.9155 | -0.2000 | 0.0000 | 0.3333 | 1.0000 |
| Matched | 1.6667 | 1.6667 | 1.1547 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| IS\_HARBOR | Unmatched | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Matched | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| Q\_AUTO | Unmatched | 1010.0000 | 1002.8000 | 312.7069 | 7.2000 | 108.0000 | 162.0000 | 372.0000 |
| Matched | 1010.0000 | 986.0000 | 317.2822 | 24.0000 | 144.0000 | 120.0000 | 168.0000 |
| Q\_NON\_AUTO | Unmatched | 1166.0000 | 1076.0000 | 374.2650 | 90.0000 | 432.0000 | 366.0000 | 474.0000 |
| Matched | 1166.0000 | 1282.0000 | 358.8147 | -116.0000 | 144.0000 | 212.0000 | 354.0000 |
| STOP\_LOCATION | Unmatched | -1.0000 | -0.8000 | 0.4140 | -0.2000 | 0.0000 | 0.3333 | 1.0000 |
| Matched | -1.0000 | -1.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

Variable:变量名

Sample:样本类型

Mean Treated: Treated组平均值

Mean Control：control组平均值

（以下数据为之前博士论文中没有展示的数据）

SD control：control组标准差

Mean Diff： 同一行Treated组和control组平均值之差

eQQ 中位数 平均值 最大值 表征的是QQ图相关的数据（分布特性相关，具体的不是很懂QQ图因此不知道怎么解释）

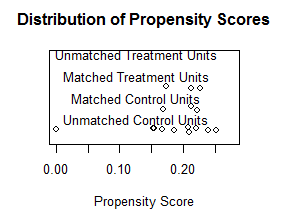
Matching effectiveness (Figure)

匹配得分及效果用图表示 由于是Treated：control=3:15，所以下图第一行 Unmatched Treatment units 下没有点，因为全部参与匹配了，

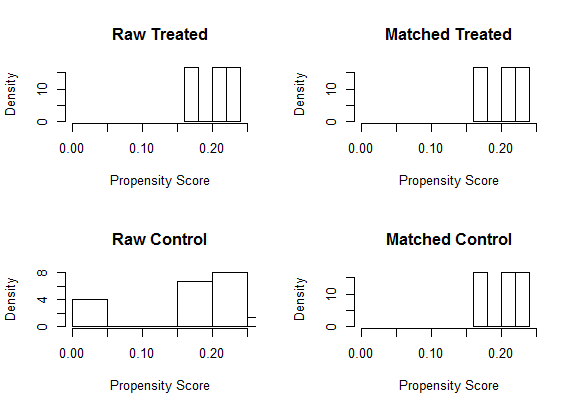
第二行Matched Treatment Unit是指用我们所用的匹配方法得到的Treated组的得分，在图中显示，

第三行 Matched control units是我们选出的三个location的倾向得分，可以跟上面第二行对比，非常相近，可见匹配效果很好

第四行是没有参加匹配的（15-3组）的PSM得分 匹配得分和效果看第二行和第三行



同理，得分和效果可以使用直方图来表示



上图Raw treated 和raw control分别是原始的treated组和control组的倾向得分，可见两者相差比较大，但是在匹配后（matched），可以看到Treated组和Matched control组的得分几乎是重合的，与上面点图那个效果一样，表征匹配的效果很好。