说明文档

**（一）文件说明**

本论文使用的数据代码压缩包命名为“2024-00551+数据”，解压之后里面有“2024-00551+**程序代码**”、“2024-00551+**日志文件**”和“2024-00551+**说明文档**”3个部分。

（1）“2024-00551+**程序代码**”里面包括6个文件，包括：

* “0\_master\_all\_do”可以执行所有的代码同时输出全文（正文+附录）的图表。
* “0\_表1”生成在线附录I图A4和图A5、表1、以及在线附录I图A6。
* “1\_表2A C+A1+A2+A3”生成图1（A）（B）（C）、表2A、在线附录II表A1的（A）（B）、表A2和表A3以及表2的第17列的回归结果。
* “1\_表2C+安慰剂检验1”生成表2第14列的回归结果以及在线附录I图A7（A）。
* “1\_表2C+安慰剂检验2”生成表2第15列的回归结果以及在线附录I图A7（B）。
* “2\_表2B+表A1---CD”生成表2（B）、图1（D）（E）（F）、在线附录II表A1（C）（D）以及表2的第18列的回归结果。
* “3\_表3”生成表3的回归结果。
* “4\_表4+表5”生成表4和表5的回归结果

（2）“2024-00551+**日志文件**”中的“table1”—“tableA3”分别对应相应表格；“fig\_1”是正文中图1的结果；“fig\_A4”、fig\_A5”、fig\_A6”、“fig\_A7\_a”、“fig\_A7\_b”分别是附录I中图A4、图A5、图A6、图A7\_a、图A7\_b。

（3）“2024-00551+**说明文档**”即是本说明文档。

**（二）数据来源**

（1）中国家庭金融调查2015（CHFS）。提供了2015年北京地区受访者的信息，包括家庭收入、户主性别、年龄、受教育年限和家庭规模等。

（2）链家网。手动抓取2010-2021年的全部交易记录以及相应的房屋信息，包括交易价格、房屋面积、楼层、户型结构、建筑类型、房屋朝向、建成年代、装修情况、建筑结构、供暖方式、产权年限、配备电梯、房屋年限和房权所属。

（3）资源环境科学数据平台。下载2020年北京市（街道乡镇）的矢量地图、数字高程数据（https://www.resdc.cn/data.aspx?DATAID=123）和GSOD气象站点数据，使用ArcGIS计算得到每个辖区的海拔、坡度、与河流的距离、平均气温和辖区面积。

（4）2015年1%人口普查数据。提供了劳动者的居住地、年龄、学历、受教育年限等信息。

（5）高德地图POI（2012）。提供了北京地区所有兴趣点的空间数据，包括学校、地铁站、政府机构和现存的世界遗产、寺庙教堂、历史遗迹纪念馆等。

（6）宋晶晶等（2017）。提供了北京市五环内在1969、1978、1990、2000和2008年的路网矢量数据。

**（三）Dofile里面使用的指标说明**

**（1）数据data1**

area\_code: 区划代码2020

south: 城南

north: 城北

geo\_x: 质心的经度

geo\_y: 质心的纬度

treat: 城南==1

distance: 辖区质心到边界的距离

NEAR\_FID: 最近边界

ln\_haiba: 海拔

podu: 坡度

ln\_dis : 与河流的距离

qiwen\_mean : 平均气温

ad\_area: 行政区面积

dist\_bd: 与边界距离(对照组取负值)

inter: treat\*dist\_bd

edu\_n: 教育机构(学校)数量

edu\_unvi: 大学数量

lnedu: ln教育机构(学校)数量的密度

lnedu\_unvi: ln大学的密度

subway: 地铁站数量

lnsubway\_d: ln地铁站密度

密度1969: ln密度1969

密度1978: ln密度1978

密度1990: ln密度1990

密度2000: ln密度2000

密度2008: ln密度2008

govern: 政府机构数量

lngov\_d: ln政府机构密度

hist\_poi: 历史准政府机构数量

lnhisgov\_d: ln历史准政府机构密度

**（2）数据house\_data**

x: 经度(房屋)

y: 纬度(房屋)

lon\_c: 经度(正阳门)

lat\_c: 纬度(正阳门)

north: 城北==1

distance : 与边界的距离

treat : false\_城南==1

price: 房屋交易价格

lnprice: ln房屋交易价格

area: 房屋面积

year: 交易年份

level: 楼层

str: 户型结构

type: 建筑类型

direction: 房屋朝向

years: 建成年代

inner: 装修情况

str\_1: 建筑结构

heat: 供暖方式

year\_d1: 产权年限

elevator: 配备电梯

year\_d2: 房屋年限

right: 房权所属

dist\_bd: 与边界的距离(对照组为负)

inter: dist\_bd\*treat

dist\_bd2: dist\_bd 2次方

dist\_bd2d: dist\_bd2次方\*treat

dist\_bd3: dist\_bd 3次方

dist\_bd3d: dist\_bd3次方\*treat

**（3）数据hh\_data**

total\_income: 家庭总收入

lnincome: ln家庭总收入

swgt: 抽样权重

hhid: 家庭ID

treat: 城南==1

dist\_bd: 与边界距离(对照组取负值)

inter: treat\*dist\_bd

dist\_bd2: dist\_bd 2次方

dist\_bd2d: dist\_bd2次方\*treat

dist\_bd3: dist\_bd 3次方

dist\_bd3d: dist\_bd3次方\*treat

age: 户主年龄

male: 户主性别

edu: 户主受教育年限

scale: 家庭规模

distance: 与边界的距离

area\_code: 区划代码2020

**（4）数据house\_placeboA和house\_placeboB中treat表示虚假的城南，相应的dist\_bd表示每个房屋与虚假边界的距离，除此之外的其他变量定义与house\_data保持一致。**

**（5）数据census15\_data**

High: 上高中及以上==1

edu\_year: 受教育年限

treat: 城南==1

dist\_bd:与边界距离(对照组取负值)

inter: treat\*dist\_bd

distance: 与边界的距离

NEAR\_FID: 最近边界#

age\_cohort: 年龄队列

area\_code : 区划代码2020

\* 分队列

group "年龄组别"，

其中：

group = 1 if age\_cohort==1 | age\_cohort==2

group = 2 if age\_cohort==3 | age\_cohort==4

group = 3 if age\_cohort==5 | age\_cohort==6

group = 4 if age\_cohort==7 | age\_cohort==8

group = 5 if age\_cohort==9 | age\_cohort==10