# 二手房房价影响因素分析

目录

[(一)单位面积房价分析 1](#_Toc3967900)

[（二）影响房屋面积的因素分析 2](#_Toc3967901)

[2.1 自变量自身分布分析 2](#_Toc3967902)

[2.2 房屋影响因素对房屋价格影响分析 3](#_Toc3967903)

[（三）建立房价预测模型 9](#_Toc3967904)

[3.1 建立模型 9](#_Toc3967905)

[3.2 预测模型 12](#_Toc3967906)

[（四）结论 13](#_Toc3967907)

# (一)单位面积房价分析

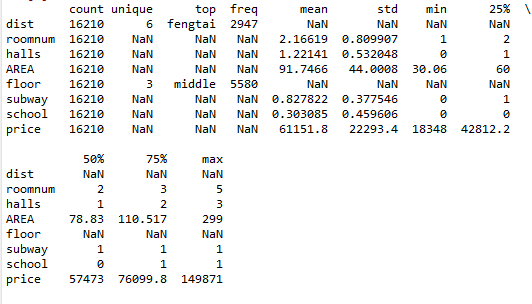


图1-1 数据基本描述

图1-1可知分析的数据量有16210个二手房源数据，其中房屋价格的因变量有dist-区域、roomnum-卧室数量、halls-客厅数量、AREA-房屋面积、floor-所在楼层、subway-是否临近地铁、school-是否是学区房。后续分析影响房屋价格的因素主要从这几个方面进行分析。

除了面积和价格是连续变量外，其余的都是名义变量。名义变量，适合使用频次统计或者百分占比进行统计描述，连续型变量：适合使用中位数，平均数，方差进行分析，分析它的趋势情况。所以后续主要从这几个方面进行分析。

# （二）影响房屋面积的因素分析

## 2.1 自变量自身分布分析

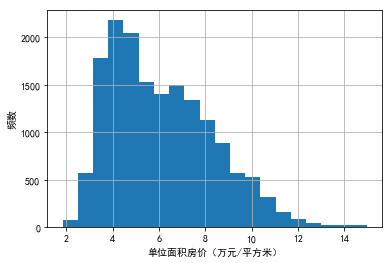


图2-1 单位面积房价直方图

图2-1中显示北京二手房的单位房屋面积房价大约在每平方米六万元左右。而且可以明显看出北京市二手房单位房屋面积价格呈现明显的右偏，为了建模的拟合度更高后续的建模前要对数据进行取对数后进行建模。

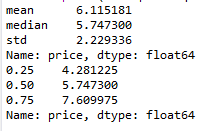


图2-2 价格的均值、中位数和标准差信息

图2-2北京二手房单位面积房价的均价为6.12万，中位数为5.7万，标准差为2.3万，推测出北京二手房的房价的高低差距会比较大。



图2-3房价最高和最低的数据

图2-3直接证实了上方的推测是正确的，北京二手房分析的数据源中单位面积房价最低位于丰台区，单位面积房价为1.9万，最高的在西城区，高达15.0万。两者的差距可谓悬殊。而且两者的房屋面积中明显单位面积房价高的比低的要小，两者的房间数和客厅都相差不大的。

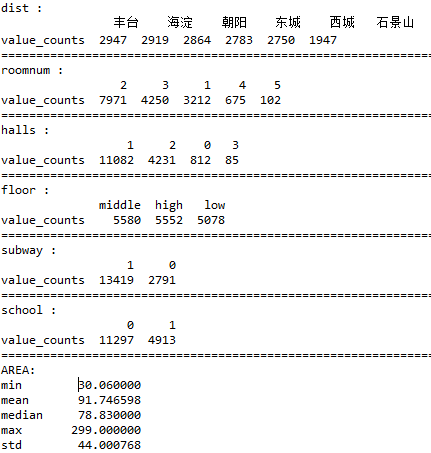


图2-4 因变量整体情况

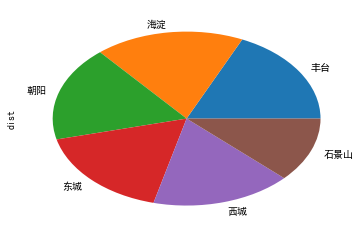


图2-5 北京各个区二手房数量柱形图

上图可看出各个区的房源数量基本相当，石景山的数量相比其它各区要少一些。但是为了后续数据建模分析单位面积房屋价格影响因素的分析，对每个区的房屋数据要进行抽样后再进行数据建模会对建模的准确性的提高更有利。

## 2.2 房屋影响因素对房屋价格影响分析

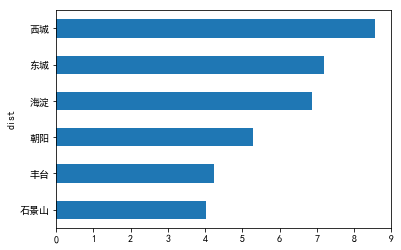


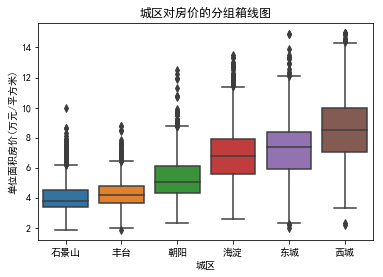
图 2-3 不同城区的单位房价面积均值情况

图2-3-1城区对房价的分组箱线图

由图2-3不同城区的单位房价面积均值降序的柱形图，可可直观地看出区域与房价的关系还是比较的密切的，不同区域的房屋均价相差比较大，初步可以得出区域是房屋价格影响因素的重要主城部分。

图2-3-1以区域的单位面积房价中位数的升序画箱型图，可以直观看出不同区域的房屋单位面积房价中位数差异明显，所以更加有力地初步证实区域会明显影响单位面积房价的推论。

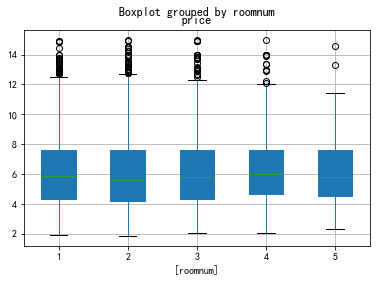


图 2-4 不同卧室数的单位面积房价差异

上图中使用卧室的数量分类画者箱型图统计，卧室数量的中位数基本相差不大，可以知道不同卧室数的单位面积房价差异不大。

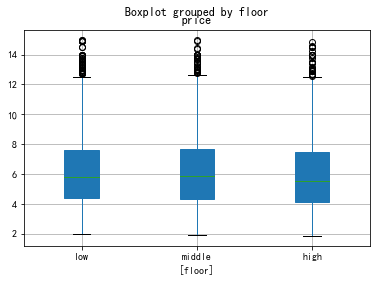


图2-5所在楼层对房价影响

上图是以楼层数分类画出所在楼层对房价影响箱型图，图中楼层数中位数相差不明显，所以不同楼层的单位房价与楼层书的差异不明显。

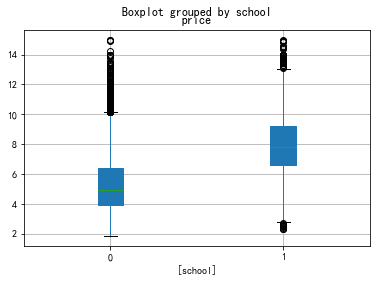


图2-6是否是学区房对房价影响

图2-6 以是否是学区房分组，画出箱型图，不是学区房的单位房价与是学区房的单位房价的中位数相比较，差距十分明显，所以学区房的房价明显比不是学区房的单位房价要高的多。

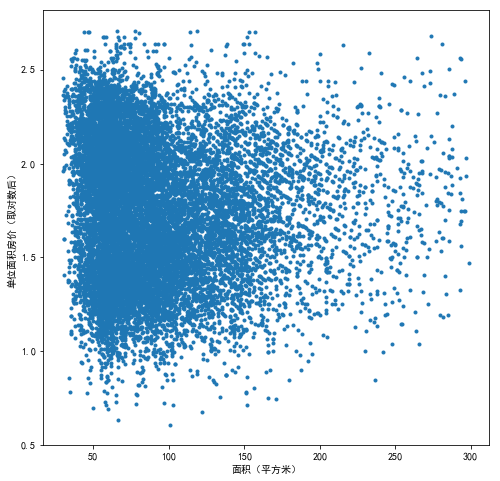


图2-7房屋面积和单位面积房价（取对数后）的散点图

对数化的房屋面积与对数化的单位面积房价存在一定的负相关（相关系 数=-0.09）。同等面积房屋的单位面积房价波动较大，尤其100平米以下的房屋

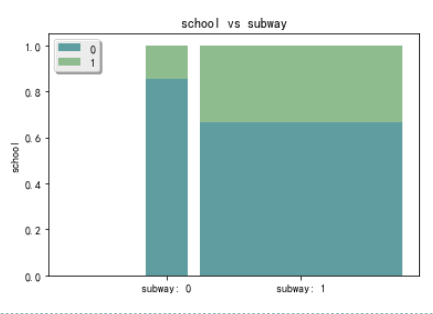


图2-8 学区房、地铁房交图

图2-8中学区房、地铁房的单位面积房价高。样本中，地铁房中的学区房比重比 非地铁房中的学区比重更大。

# （三）建立房价预测模型

## 3.1 建立模型

由于原始样本量太大，无法使用基于P值的构建模型的方案，因此按照区进行分层抽样。按照区分层，每个区抽取400个样本。社会科学取值偏大,二手房价社会科学问题，因此取0.05较为合理,经过多次的随机抽样后发现，dist的P值为:0.0000说明区域对房价的影响时很直接的。roomnum的P值为:0.8379说明卧室数对单位面积房价影响不大，因此建模时可舍弃。halls的P值为:0.2372客厅数与单位面积房价有轻微影响。floor的P值为:0.0702不同楼层的单位面积房价差异不明显。subway的P值为:0.0000地铁房单价高，school的P值为:0.0000学区房的房价也高。

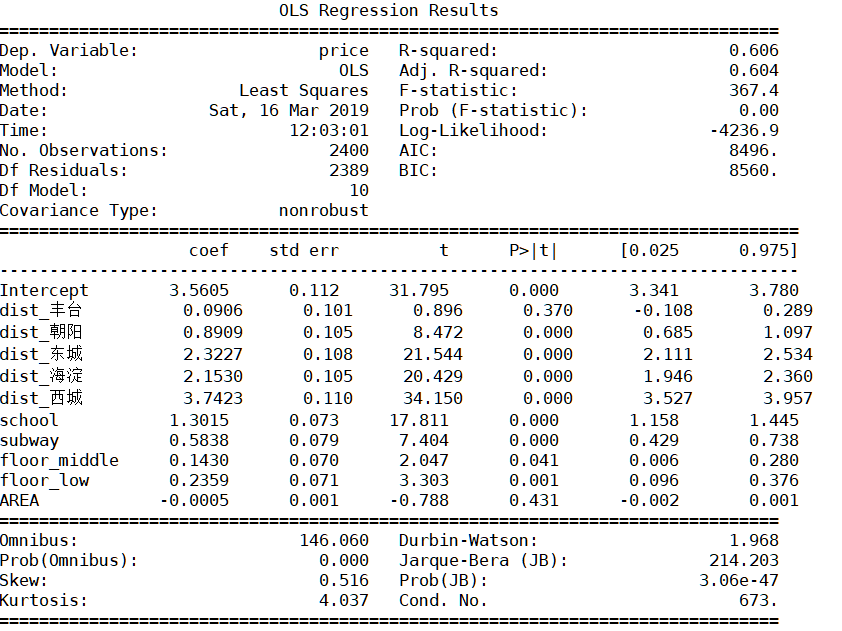


图3-1线性回归模型

图3-1控制其他因素不变时 城区：石景山区单位面积房价最低，西城区单位面积 房价最高，比石景山区每平米平均高出3.67万元 学区房比非学区房单位面积房价平均高出1.20万元 地铁房比非地铁房单位面积房价平均高出5633元 高层房屋单位面积房价最低，其次是中层，低层房屋 面积房价最高 房屋面积的增加会带来单位面积房价的降低.

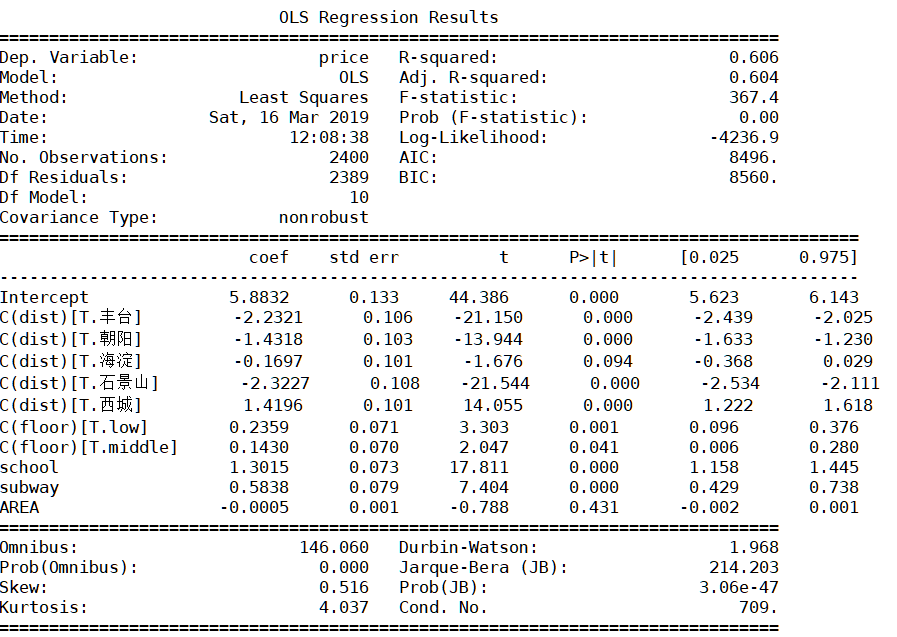


图3-2

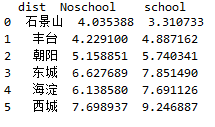


图3-3不同区域学区房和非学区房单价

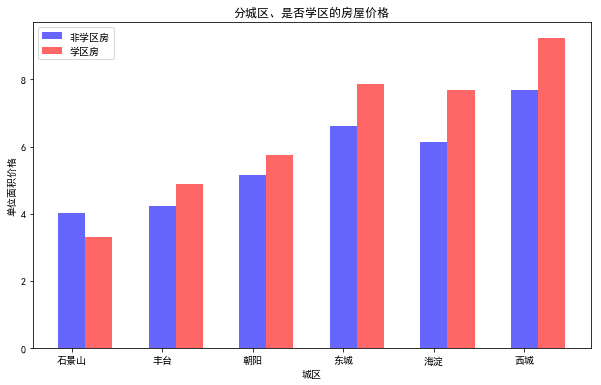


图3-4不同区域学区房和非学区房单价

图3-3和3-4可以直观地观察到不同区域是否是学区房对单位面积房屋价格的影响不尽相同。石景山区的非学区房要比学区房的单位面积房价要高几千。其中朝阳区的差价不是特别大，而西城区的差价就十分明显相差1.6万左右。

## 3.2 预测模型

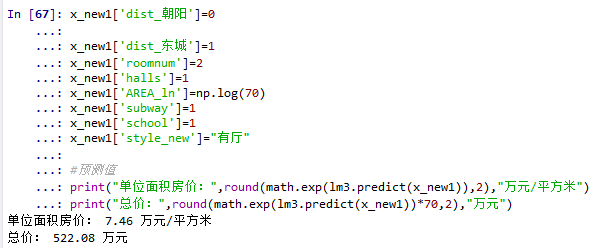


图3-2-1

假想：一家三口，为了孩子在东城区上学，想 买一套邻近地铁的两居室，面积是70平方米， 中层楼层 λ 根据交互模型，预测的单位面积房价是7.86万 元/平方米，总价550.37万元

# （四）结论

影响北京市二手房单位面积房价的主要因素 ⎫ 区位因素：城区、地铁、学区，内部因素：房屋面积、楼层 λ 使用对数线性模型，可以克服数据中存在的异方差 λ 使用交互模型，能带来更好的模型解读 ⎫ “学区优势”对各城区单位面积房价的影响有所区别。