

邯郸市二手房房价影响因素分析

周鏊 赵芳卉 陈琳

(指导教师: 张艳萍)

河北工程大学

一、研究背景

自 1998 年国家取消福利分房,我国房地产行业开启了市场化进程。经过 20 年的快速发展,房地产行业逐渐成为国民经济新的增长点,为中国经济的快速增长做出了贡献^[1-3]。据国家统计局数据,2000 年我国房地产开发企业的营业利润只有 73.3 亿元,而到 2016 年利润已达到 8673.2 亿元^{[4][5]}。1999 年至 2018 年间,我国商品房平均销售价格从 2051 元/平方米迅速增至 8737 元/平方米。北京、上海、广州、深圳、苏州、南京等一二线城市的房价涨幅更大^[7]。房地产行业繁荣发展的同时,也带来了一系列的社会问题,因而备受关注。

为了解决房价持续增长的问题,中央和地方出台了多项调控政策。2014 年 3 月,国务院发布《国务院办公厅关于继续做好房地产市场调控工作的通知》,2015 年 9 月,中国人民银行、中国银行业监督管理委员会发布《关于进一步完善差别化住房信贷政策有关问题的通知》,2016 年 8 月,住房城乡建设部等部门发布《关于加强房地产中介管理促进行业健康发展的意见》^[8]。然而这些措施在稳定房价方面收效甚微。伴随着房价的飞速增长,购买能力相对较低的消费者选择更多地关注到二手房市场,二手房市场迎来了新的春天^[9-11]。以北京市为例,二手房市场已经到来,2018 年北京二手房成交量占房屋总成交量的 86%,而新建商品房成交量仅占 14%。

二手房对补充住宅房源,增加住房选择,减轻住房压力,提升人民生活水平和幸福感有重要的意义,因而对二手房的研究十分有必要。

二、研究目的

本文以河北省邯郸市二手房市场为例,从二手房的实际环境方面,分析影响二手房房价的多个因素,通过运用定性分析和定量分析工具,结合描述统计和推断统计的方法,选取影响二手房价的主要因素,以期达到以下目的:

1. 判断所选的变量对二手房价是否有影响,并分析各个因素对二手房房价

的影响程度。

2. 从邯郸市实际情况出发，分析出邯郸市二手房房价的影响机制，并构建模型，依据案例数据计算出二手房的合理价格。

3. 通过该研究，一方面为购房者的购买决策、售房者的定价决策、投资者投资决策提供依据；另一方面为政府的政策制定提供建议及参考依据。

三、数据来源

本文研究总体为邯郸市二手房房源，基于样本准确性、全面性、代表性和抽样过程的可靠性，本文以分层抽样为抽样方法，通过搜房网进行区域筛选，共收集 300 个邯郸市二手房样本，每个样本各有 8 个指标。

本文采用的简单随机抽样方法如下：按照行政区域的划分，从邯郸市下设的行政区中分别随机抽取 15 个样本，最终共得到 300 个样本。

四、指标选取

影响二手房的因素有很多，本文在参考了大量文献基础上，详细咨询了从事房地产工作的专业人士，并最终选取了以下指标为解释变量：

1. 房子年龄（连续型变量）：自房屋竣工验收合格交付使用之日起到 2019 年 6 月 1 日为止的时间
2. 建筑面积（连续型变量）：二手房的建筑面积
3. 楼层（连续型变量）：二手房所在楼房的楼层总数
4. 户型（离散型变量）：包括二室、三室、四室、五室及以上
5. 装修程度（离散型变量）：豪华装修、简单装修、精装修、毛坯房
6. 所在区（离散型变量）：本文中共包括邯郸市的行政区域：邯山区、丛台区、复兴区、峰峰矿区、肥乡区、永年区、武安市、临漳县、成安县、大名县、涉县、磁县、广平县、馆陶县、魏县、曲周县、高开区、冀南新区
7. 房屋朝向（离散型变量）：东、西、南、北、西东、南北
8. 产权性质（离散型变量）：商品房、住宅房、商住两用、经济适用房

五、统计描述

（一）房价分析——房价大概在 7000 元/平方米上下波动

对邯郸市二手房房价进行分析，发现二手房房价大致呈现正偏态分布，其偏度约为 0.51，即高峰向左偏移，如图 1 所示，从图中可知，约六成二手房的房价均处于 4000 元/平方米至 10000 元/平方米之间。通过计算可得，邯郸市二手房房价的均值为 7472.968 元/平方米，中位数为 7180.642。由此，我们可推断邯郸市二手房房价大致在 7000 元/平方米上下波动，为探究影响邯郸市二手房房价的关键因素，下文我们将对收集的 8 个变量与房价进行交叉分析。

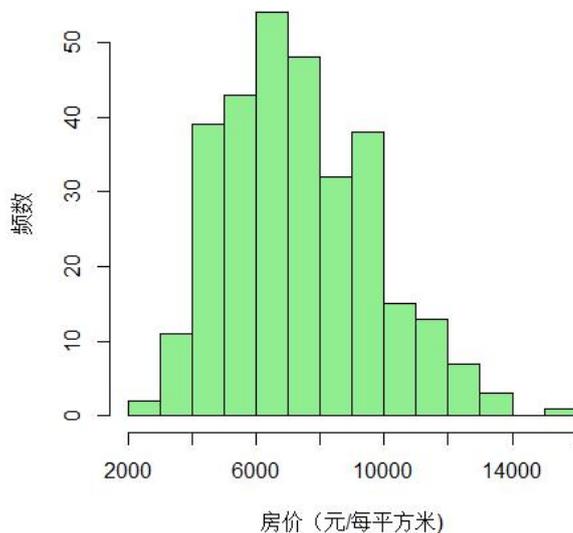


图 1 邯郸市二手房房价分布图

（二）行政区域对房价的影响分析

从图 2 的区域平均价格箱线图可比较发现，在邯郸市下设的行政区域中，涉县的二手房价格总体水平最高，二手房价格水平排名前三的分别是涉县、高开区、邯山区，二手房价格排名倒数前三的分别是大名县、永年区、魏县。总体来看旅游城市和开发区的房价要普遍高于经济实力相对较弱的地区，这显然符合区域对房价有影响的假设。同时，从箱线图上可以看出，丛台区、峰峰矿区、广平县、临漳县的二手房房价变化幅度较大，产生这种结果的可能是区域面积较大，发展水平不均衡等。

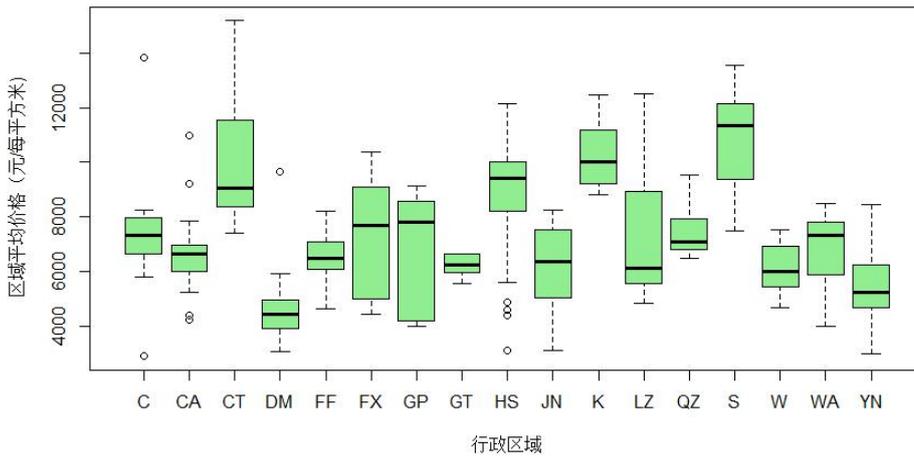


图2 邯郸市各行政区域所对应的房价箱线图

注：C表示磁县、CA表示成安县、CT表示丛台区、DM表示大名县、FF表示峰峰矿区、FX表示复兴区、GP表示广平县、GT表示馆陶县、HS表示邯山区、JN表示冀南新区、K表示高开区、LZ表示临漳县、QZ表示曲周县、S表示涉县、W表示魏县、YN表示永年区。

(三) 楼龄对房价的影响分析

本文抽取的二手房源样本，楼龄高低不一，低至不足一年，高至39年，由图3中的二手房楼龄分布直方图可以发现：邯郸市二手房的楼龄大致呈右偏分布，楼龄为3年至9年的二手房约半数。

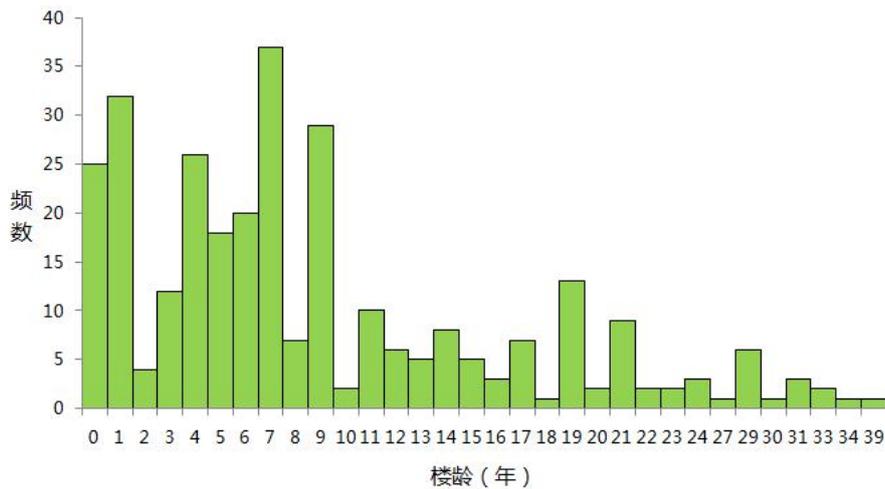


图3 邯郸市二手房房龄分布直方图

（四）户型对房价的影响分析

为研究户型对邯郸二手房房价的影响，本文绘制了不同户型所对应的房价的箱线图，如图 4 所示，五市及以上的房源最少，由于这些类型的住宅房只有一套，因此其箱线图显示为一条直线。二室的平均房价最高，超过 8000 元/平方米，三室的平均房价最低，约为 7000 元/平方米。产生这种结果的原因是三室的房源比较多，房源供应量充足，所以价格比较低；一室二室的房源基本上是应时代下的单身男女的需求而建的，另因住房比较新，装修比较豪华等因素，故一室二室的平均房价高于三室。

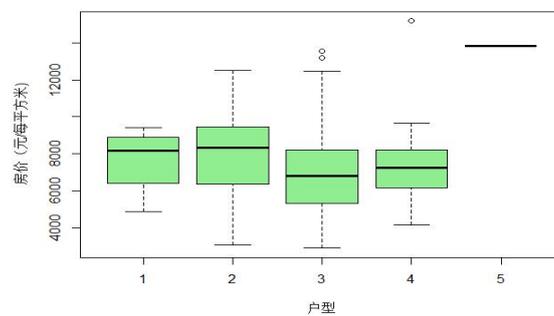


图 4 户型对应的房价箱线图

注：1 表示一室、2 表示两室、3 表示三室、4 表示四室、5 表示五室及以上

（五）楼层对房价的影响因素分析

楼层的高低与基础消防要求、结构要求和技术手段等因素息息相关，从而对房价产生影响，因此研究楼层对房价的影响是十分必要的。对数据进行分析，得到楼层频数分布图⁹，即图 5。从图 5 可发现，邯郸市二手房源中，4 到 7 层高的楼源最多，基本占了本案例中所有二手房源的一半。由此可知，大部分的二手房都位于 10 层以下。我们也可以同时推断出：大部分的二手房都是有一定楼龄的旧楼房。

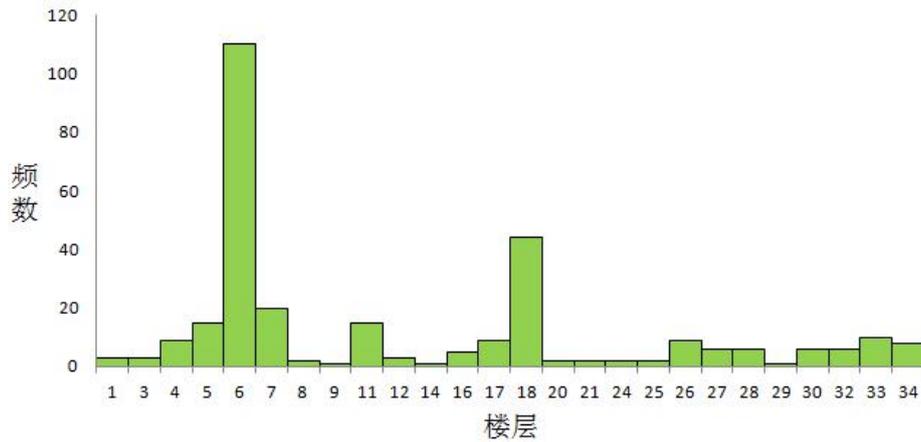


图5 楼层分布直方图

(六) 装修程度对房价的影响因素分析

根据房子的装修程度，二手房可简单分为“毛坯房”、“简单装修”、“精装修”、“豪华装修”四种装修程度的房子。从图6中可以看出，毛坯房的房价最低，简单装修的变化程度最大，精装修的上下四分位数及中位数都略低于简单装修，豪华装修虽然样本较少，但从图中可看出，豪华装修房价的上下四分位数和中位数都是最高的，即基本符合装修程度越好，房价越高的事实。

为何精装修的四分位数参数均略低于简单装修？为分析这一反常现象，本案例小组首先分别分析了精装修和简单装修中的建筑面积这一参数，发现简单装修的二手房平均建筑面积为 94.82m²，精装修的二手房平均建筑面积为 110.31m²，即建筑面积无法解释这一反常现象。

然后本案例小组分别分析了简单装修和精装修中的行政区域分布，发现在简单装修房源中发展较好的两个区中丛台区占 13 个，邯山区 16 个，其他 4 个区 25 个，而简单装修的二手房房源共有 93 个，即发展较好的区的二手房房源占简单装修的房源一半以上；精装修房源中发展较好的两个区丛台区占 3 个，邯山区占 7 个，其他 4 个区占 28 个，而精装修的二手房房源共有 117 个，即发展较好的区的二手房房源占精装修的房源 41%左右。

综上所述，出现这种反常情况的原因很大程度上跟行政区域有关。

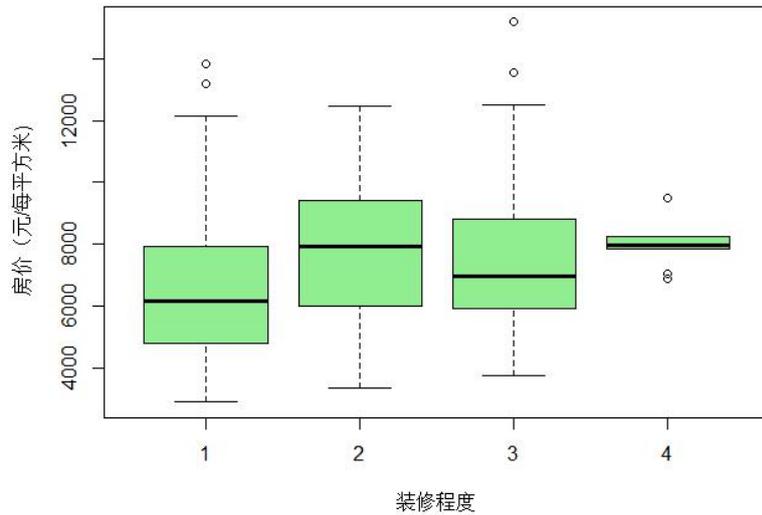


图 6 邯郸市装修程度箱线图

注：1 “毛坯”、2 “简单装修”、3 “精装修”、4 “豪华装修”

六、统计建模——构建二手房估价模型

通过以上的描述性统计分析，我们可以得出：以上因素确实与二手房价存在相关性。考虑到这些因素中有部分是离散型变量，本案例小组将建立多因素协方差模型进行分析。此外需要注意的是，前面一节提到的房价为房子单价，本节考虑到总价与面积的联系，用到的房价指数为房子总价，下面分别建立多因素协方差分析。

（一）不考虑交互作用的多因素协方差分析模型

将上文分析的 8 个自变量引入回归模型，构建协方差分析全模型如下：

$$\text{房子总价} = \beta_0 + \beta_1 \times \text{楼龄} + \beta_2 \times \text{面积} + \beta_3 \times \text{楼层} + \beta_4 \times \text{户型} + \beta_5 \times \text{装修程度} + \beta_6 \times \text{行政区域} + \beta_7 \times \text{朝向} + \beta_8 \times \text{产权性质} + \varepsilon$$

——模型（1）

通过 R 统计软件对模型进行协方差分析，分析结果如表 1 所示。

表 1 模型 (1) 协方差分析结果

	Df	Sum sq	Mean sq	F value	Pr(>F)
楼龄	1	11073	11073	39.2182	1.47E-09
面积	1	98552	98552	349.0428	2.20E-16
楼层	1	23128	23128	81.912	2.20E-16
户型	4	44855	11214	39.7159	2.20E-16
装修程度	1	5872	5872	20.3583	9.54E-06
行政区域	16	47266	2976	10.3185	2.2e-16
朝向	4	1108	277	0.9603	0.4297
产权性质	3	1509	503	1.7441	0.1582

通过此表我们可以发现，朝向和产权性质没有通过显著性检验，我们把上述两个变量从模型中剔除。

(二) 不考虑交互作用的多因素协方差修正模型

将朝向和产权性质这 3 个没有通过参数显著性检验的变量从全模型中剔除后，得到模型 (2)：

$$\text{房子总价} = \beta_0 + \beta_1 \times \text{楼龄} + \beta_2 \times \text{面积} + \beta_3 \times \text{楼层} + \beta_4 \times \text{户型} + \beta_5 \times \text{装修程度} + \beta_6 \times \text{行政区域} + \varepsilon$$

——模型 (2)

通过 R 统计软件对模型进行协方差分析，分析结果如表 2 所示。

表 2 模型 (2) 协方差分析结果

	Df	Sum sq	Mean sq	F value	Pr(>F)
楼龄	1	11073	11073	38.108	2.333e-09
面积	1	98552	98552	339.162	2.2e-16
楼层	1	23128	23128	79.593	2.2e-16
户型	4	44855	11214	38.592	2.2e-16
装修程度	1	5872	5872	20.209	1.016e-05
行政区域	16	47266	2976	10.243	2.2e-16
Multiple R-squared: 0.7389		Adjusted R-squared: 0.7166			

由表 2 结果显示, 在剔除模型 (1) 中的不显著变量并对模型 (2) 进行协方差分析后, 模型 (2) 中的变量均通过了显著检验。可决系数 $R^2=0.7389$, 调整后的 $R^2=0.7116$, 由此可见, 模型 (2) 的拟合优度较好。通过对模型 (2) 中的参数进行估计, 得到结果如表 3 所示。

表 3 模型 (2) 参数估计

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-30.10637	11.10246	-2.712	0.007107
楼龄 s	0.37661	0.19341	1.947	0.052499
面积	0.64641	0.06066	10.657	2e-16
楼层	0.47750	0.14772	3.233	0.001373
户型 2	13.20477	9.18584	1.438	0.151685
户型 3	17.43759	9.60436	1.816	0.070499
户型 4	31.27750	11.35053	2.756	0.006241
户型 5	214.64409	214.64409	9.251	2e-16
装修程度	5.58694	1.29999	4.298	2.38e-05
行政区域 CA	-7.22987	5.76955	-1.253	0.211208
行政区域 CT	20.33325	6.21738	3.270	0.001208
行政区域 DM	-27.23460	6.40437	-4.253	2.88e-05
行政区域 FF	-8.17066	6.57855	-1.242	0.215267
行政区域 FX	-1.71322	5.59024	-0.306	0.759476
行政区域 GP	-7.14736	8.10961	-0.881	0.378885
行政区域 GT	-8.41147	9.09580	-0.925	0.355882
行政区域 HS	12.01458	5.94859	2.020	0.044360
行政区域 JN	-8.14607	5.80971	-1.402	0.161975
行政区域 K	24.50943	6.73500	3.639	0.000326
行政区域 LZ	-4.58233	6.55307	-0.699	0.484965
行政区域 QZ	2.05246	6.42600	0.319	0.749660
行政区域 S	32.45951	6.59467	4.922	1.46e-06
行政区域 W	-12.29834	6.56270	-1.874	0.061971
行政区域 WA	-8.83280	6.57161	-1.344	0.180005
行政区域 YN	-23.38886	5.82841	-4.013	7.70e-05

对模型 (2) 的协方差分析和参数估计, 得出房价的定价方差如下:

$$\text{房子总价} = -30.10637 + 0.37661 \times \text{楼龄} + 0.64641 \times \text{面积} + 0.47750 \times \text{楼层} \\ + \beta_4 \times \text{户型} + 5.58694 \times \text{装修程度} + \beta_6 \times \text{行政区域}$$

——模型 (3)

通过模型(3)可以得出如下结论:

1) 楼龄对房总价会有正影响。一般在其他因素不变的情况下,楼龄越长,代表着房屋的折旧越高,剩余可使用年限也就越短,房价越低⁰。出现与常识不符的情况的可能性原因是跟邯郸的具体情况有关。邯郸市6月二手房均价9772元/m²,但是各个区的二手房均价却相差很大,例如:永年4872元/m²,丛台10864元/m²,广平2417元/m²。其中丛台区为发展较好的区,但近几年新建小区比较少,把楼龄对房价的负影响拉高成了正影响。

2) 建筑面积和房子总价正相关,即在其他因素相同的情况下,面积越大,总房价越高,在其他因素不变的条件下,面积每增加1平方米,房子总价会增加0.64641万元。

3) 楼层和房子总价呈现正相关关系,即在其他因素不变的情况下,楼层越高,房子总价越高,出现这种情况的原因是高楼层采光好且噪音较少⁰,即当其他因素相同情况下,楼房的整体高度每提高一层,房子总价大约提高0.47750万元。

4) 装修程度跟房子总价正相关,即在其他因素不变的情况下,简单装修比毛坯房、精装修比简单装修的房、豪华装修比精装修的房子总价都大约高6万元左右。

(三) 考虑交互作用的多因素协方差分析模型

在上述多因素协方差分析模型的建模过程中,我们发现有些因素对房价影响是不显著的,但不能说明这些因素对目标变量无影响,因为这些不显著的因素可能与其他因素存在交互作用,进而对房价产生影响。因此将不显著因素直接删除,然后进行回归,所得到的模型可能不是最优的。为优化模型,接下来将会考虑自变量之间的交互作用对目标变量的影响,将其引入模型后再进行建模。

通过参考文献以及以往研究者对房价的研究并结合实际,本文提出了变量间可能存在的两种交互作用。

第一，“朝向”与“户型”有交互作用。设计师在设计时，户型会对房子的南北朝向产生一定的影响。因此可以将两者结合起来进一步细分户型这一变量，以研究两者对房价的影响。

第二，“产权性质”与“行政区域”有交互作用。邯郸市有些行政区域商住两用房较多，有些行政区域住宅房更多，个别区域经济适用房较多，而另一些区域商品房更多。因此可以将两者结合起来进一步细分行政这一变量，以研究两者对房价的影响。

为验证上述假设是否成立，本文将两种可能存在的交互作用引入模型进行恢复，首先进行全模型回归，得到模型（4）：

$$\text{房子总价} = \beta_0 + \beta_1 \times \text{楼龄} + \beta_2 \times \text{面积} + \beta_3 \times \text{楼层} + \beta_4 \times \text{户型} + \beta_5 \times \text{装修程度} + \beta_6 \times \text{行政区域} + \beta_7 \times \text{朝向} + \beta_8 \times \text{产权性质} + \beta_9 \times \text{朝向} \times \text{户型} + \beta_{10} \times \text{产权性质} \times \text{行政区域} + \varepsilon$$

——模型（4）

对数据处理后，在 R 中得到协方差分析结果，如表 4 所示。

表 4 模型（4）协方差分析结果

	Df	Sum sq	Mean sq	F value	Pr(>F)
楼龄	1	11073	11073	40.1849	9.656e-10
面积	1	98552	98552	357.6470	2.2e-16
楼层	1	23128	23128	83.9311	2.2e-16
户型	4	44855	11214	40.6949	2.2e-16
装修程度	1	5872	5872	21.3107	6.045e-06
行政区域	16	47622	2976	10.8012	2.2e-16
朝向·	4	1108	277	1.0052	0.4052121
产权性质	3	1509	503	1.8257	0.1427464
户型*朝向·	1	1	1	0.0034	0.9536333
行政区域*产权性质	3	4633	1544	5.6047	0.0009652

结果显示考虑交互作用后，变量房屋朝向与户型的交互作用对目标变量（房总价）仍无显著影响，但原先对房价无显著影响的产权性质，在模型中加入行政区域的交互作用后，协方差分析的结果显示交互作用对房价有显著影响，同时也说明产权性质这个变量不应该被剔除。

(四) 考虑交互作用的多因素协方差分析修正模型

将通过显著性检验的变量从模型中剔除并加入交互影响因素后, 得到回归模型如下:

$$\begin{aligned} \text{房子总价} = & \beta_0 + \beta_1 \times \text{楼龄} + \beta_2 \times \text{面积} + \beta_3 \times \text{楼层} + \beta_4 \times \text{户型} + \beta_5 \times \text{装修程度} \\ & + \beta_6 \times \text{行政区域} + \beta_7 \times \text{产权性质} + \beta_8 \times \text{产权性质} \times \text{行政区域} + \varepsilon \end{aligned}$$

——模型 (5)

在 R 中得到协方差分析结果, 如表 5 所示:

表 5 模型 (5) 协方差分析结果

	Df	Sum sq	Mean sq	F value	Pr(>F)
楼龄	1	11073	11073	40.4352	8.473e-10
面积	1	98552	98552	3598746	2.2e-16
楼层	1	23128	23128	84.4539	2.2e-16
户型	4	44855	11214	40.9484	2.2e-16
装修程度	1	5872	5872	21.4434	5.635e-06
行政区域	16	47622	2976	10.8685	2.2e-16
产权性质	3	1656	552	2.0154	0.112032
行政区域*产权性质	4	4961	1240	4.5286	0.001474

通过表 5 我们可以看出, 各变量均通过显著性检验, 即每个变量对房价都有显著影响。结合回归分析结果, 得到以下结论:

- (1) 户型对房价的影响是正向的, 即户型越大, 房价越高。
- (2) 区域与房价正相关, 行政区域发展越好, 房价越高。

本文所构建的模型 (5) 即为二手房房价的估价模型。

七、基于 Lasso 方法的房价影响因素选择

(一) Lasso 介绍

Lasso 全称 Least absolute shrinkage and selection operator, 是由 Robert Tibshirani 在 1996 年提出的一种能够实现指标集合精简的估计方法。Lasso 通过

把系数的绝对值控制在一个常数范围内使得变量的残差平方和最小,由于这种方法具有压缩的特点,使一些估计系数变为 0,因此得出了有解释力的模型。其具体形式为:

$$\min_{\beta} \left\{ (y - x\beta)^2 + \sum_j \lambda (|\beta_j|) \right\} \quad (6)$$

其中, $(y - x\beta)^2$ 为平方损失函数, $\sum_j \lambda (|\beta_j|)$ 为对系数变量 β 的惩罚,由于惩罚比例都是 λ ,因此公式 (2) 可变形为:

$$\min_{\beta} \left\{ (y - x\beta)^2 + \lambda \sum_j |\beta_j| \right\} \quad (7)$$

因此, Lasso 方法是相对每一个 $\lambda \in (0, \infty)$, 计算向量使得上述函数最小化。其原理如下:

当 $\lambda = 0$ 时, 上述函数变为最小二乘, 向量 β 是真实的估计

当 $\lambda = \infty$ 时, 使上述式子最小化的向量 $\beta = 0$, 即对于每一个元素 $\beta_j = 0$

当调节参数 λ 从 ∞ 趋向于 0 时, 从 0 到变为非 0 的向量 β 也会越来越多, 而比较有影响的元素也会首先变为非 0。系数变为非 0, 即相对应的解释变量对因变量具有较重要的解释作用, 于是达到变量选择的目的。

虽然最小二乘法 OLS 具有很多的优点, 但仍不能满足某些情况下的数据分析要求。首先是 OLS 只能筛选出对因变量存在影响的变量, 但不能对各变量的重要性作出判断, 而 Lasso 方法比较好地解决了这个问题。其次尽管 OLS 是无偏估计, 但它的方差在自变量线性相关程度高时通常较大, Lasso 是通过压缩一些系数到 0 来改进预测精度解决了这个问题。

基于以上最小二乘 OLS 的不足之处的考虑, 为将房价定价模型更优, 本节将采用 Lasso 方法建立 Lasso 回归模型。

(二) 数据处理

由于 Lasso 回归模型中只能对数值型变量进行回归分析, 而行政区域这个变量不为数值型, 考虑到每个区域的发展情况不同, 再结合从咨询二手房房地产专业人员获得的信息, 我们把行政区域的数据进行转化, 数据处理方法如表 6 所示。

表 6 区域的数据处理

变量	对应编码	变量	
CT (丛台区)	1	CA (成安县)	10
FX (复兴区)	2	W (魏县)	11
HS (邯山区)	3	DM (大名县)	12
JN (冀南新区)	4	QZ (曲周县)	13
K (经济开发区)	5	LZ (临漳县)	14
WA(武安市)	6	C (磁县)	15
YN (永年区)	7	GT (馆陶县)	16
S (涉县)	8	GP (广平县)	17
FF (峰峰矿区)	9		

(三) 变量选择

根据前面几节构建的协方差分析模型的结果，我们用 Lasso 方法对房价可能存在的显著性变量进行选择排序，使用 R 并引入相关编程。输出的结果为 Lasso 方法的解路径（未对数据标准化），解路径如图 7 所示。

由图 7 可以看出，影响房价的因素中，按影响程度大小排序依次如下：
4（户型）6（行政区域）5（装修程度）2（面积）3（楼层）1（楼龄）

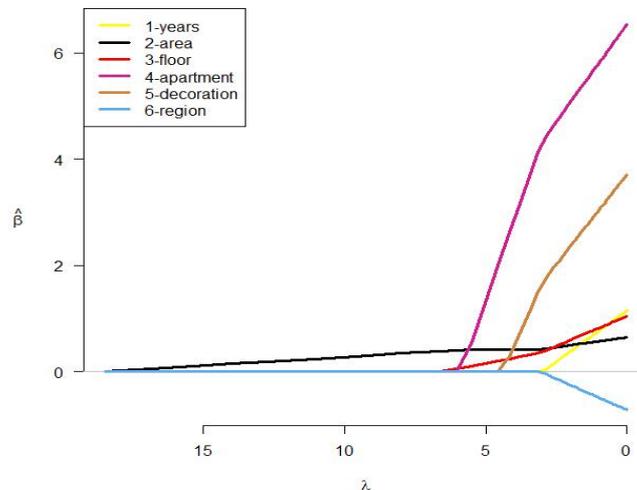


图 7 Lasso 方法的图解路径

(四) 参数估计与解释

我们对以上变量通过 Lasso 回归进行参数估计，得到估计结果如下。

表 7 Lasso 回归参数估计结果

变量	估计值	变量	估计值
(Intercept)	-36.6084426	户型	6.5254779
楼龄 s	1.1446938	装修程度	3.7010886
面积	0.6458871	行政区域	-0.7105217
楼层	1.0434736		

通过表 7 可以看出，参数估计都为非零常数，根据 Lasso 回归，这些系数都是对房价有影响的，从中可得以下结论：

- (1) 二手房楼龄和房价成正比，这个系数的反常原因在上节已提，此节不再赘述。
- (2) 房屋面积与房价正相关，根据回归结果显示，面积每提高 1m^2 ，房总价会增长 0.6458871 万元。
- (3) 楼层总数与房价正相关，根据回归结果显示，总楼层每增加 1 层，房总价会增长 1.0434736 万元。
- (4) 户型与房价为正相关关系，每增加一室，房总价会增加 6.5254779 万元。
- (5) 装修程度与房价也呈现正相关关系，即在其他条件不变的情况下，每上升一个装修程度，房总价增加 3.7010886 万元。

八、研究结论

(一) 影响二手房房价的显著因素

通过前面几节的描述统计和模型构建，我们得知邯郸市二手房的房价大致在 7000 元/ m^2 上下浮动，影响二手房房价的显著因素包括如下 7 个：

- (1) 楼龄：楼龄对房总价呈正比关系。这是因为老区发展较好，虽然楼龄长，但是一定程度上比县里的楼龄低的房子价格高。
- (2) 面积：在其他因素不变的情况下，建筑面积越大，房子相对越舒适，房总价

会相应越高。

- (3) 楼层数：楼层数和房总价正相关，即在其他因素不变的情况下，总楼层越高，房总价也会越高。
- (4) 户型：五室比四室、四室比三室、三室比二室的总房价高。
- (5) 装修程度：在其他因素相同的情况下，每提高一个装修程度，总房价就会越高。
- (6) 行政区域：不同的行政区域房价会不一样，就邯郸市而言，综合发展程度越高的行政区域，总房价越高。
- (7) 产权性质*行政区域：在某些行政区域，产权性质为商住两用的房子比其他产权性质的房子总价要低。

（二）二手房估价模型

本案例组构建的二手房估价模型如模型（8）所示：

$$\text{房子总价} = \beta_0 + \beta_1 \times \text{楼龄} + \beta_2 \times \text{面积} + \beta_3 \times \text{楼层} + \beta_4 \times \text{户型} + \beta_5 \times \text{装修程度} \\ + \beta_6 \times \text{行政区域} + \beta_7 \times \text{产权性质} + \beta_8 \times \text{产权性质} \times \text{行政区域} + \varepsilon$$

——模型（8）

根据模型（8）的估计结果公式（1），我们可以代入相关指标数据估算其二手房的价格，模型（8）的估计结果详见附录。

（三）对总房价有影响的影响因素

按照其重要性排序，依次如表 8 所示：

表 8 对房价有影响的因素的重要性排序

影响因素	重要性排名	影响因素	重要性排名
户型	1	面积	4
区域	2	楼层	5
装修程度	3	楼龄	6

九、建议

（一）给购房者的建议

对于计划在邯郸市买房的购房者，本案例小组结合参考文献资料、所研究结论并结合实际问题的应用，给出以下建议：

- (1) 购房前应该实地考察几次，以免买下有缺陷的住房而遗憾；
- (2) 注重二手房所处环境，安全第一；
- (3) 为了避免噪音的影响和方便出行，建议不要选路边的楼房或者楼顶的住房；
- (4) 细看合同条款，挑选专业并值得信任的中介。在中介的帮助下，购房者对整个过程需要了解 and 注意的事项能够心中有数，并结合本案例估算价格做辅助衡量标准，避免被中介舞蹈。

（二）给售房者的建议

对于想让自己的二手房能卖出去且能卖出好价钱的售房者，本案例小组给出以下建议：

- (1) 同购房者一样，售房者也需要细看合同条款，挑选专业并值得信任的中介。在中介的帮助下，购房者对整个过程需要了解 and 注意的事项能够心中有数，并结合本案例估算价格做辅助衡量标准，避免被中介误导。
- (2) 售房者可考察中介结构营业执照、房地产经纪从业人员资格证书、房地产经纪机构（或分支机构）备案证书等以便熟悉所选中介机构。

（三）给投资者的建议

对于想要投资风险最小和利润最大的投资者，本案例小组给出以下建议：
投资发展行政区域比较好、楼龄较低、建筑面积较大、户型较大、总楼层较高、装修程度较好的房子。

（四）给政府的建议

针对二手房这一具有前景的产业，本案例小组结合研究成果及参考资料为政府提供以下建议。

(1) 建立严格的监督管理机制。目前邯郸市政府对于二手房的监管管理机制很弱，二手房依旧存在着严重的产权问题和质量问题，降低了二手房交易的信誉，给交易带来了很大麻烦，影响了市场的健康发展。在国外，法律体系已经完善，同时也建立了个人信用制度，我们应该积极学习国外成熟的二手房交易体系并结合中国国情，建立适合我国的监督管理机制，保证二手房顺利进行交易。

(2) 中介是连接卖方和买方的纽带，是房产信息的主要来源，也是二手房市场的推动器，在二手房市场中扮演着重要的角色。政府部门应积极培育中介市场，严格规范行业，提高中介从业人员的素质，树立行业道德意识，加强对行业的监督管理，引导中介机构健康发展。

参考文献

- [1] 张岚,钱夕元,阮彤.基于 Lasso 的慢性心衰病人中西医疗效对比研究[J].中国卫生统计,2019,36(03):365-367.
- [2] 谭勇,谢林柏,冯宏伟,温子腾.基于 LASSO 回归的红外火焰探测器的设计与实现[J].激光与红外,2019,49(06):720-724.
- [3] 张平平,芮斌.二手房市场现状及存在问题分析——以安徽省芜湖市为例[J].中国房地产,2019(18):28-33.
- [4] 黄明宇,夏典.合肥市二手房价多元线性回归预测模型[J].合作经济与科技,2019(09):80-82.
- [5] 王琼.西安市二手房市场价格现状及原因分析[J].经济师,2019(04):176-177.
潘添翼,贾德铮.上海二手房房价影响因素[J].中国市场,2019(05):29-32.
- [6] 梁娜飞,梁丽娜,张慧慧.西安市房地产发展现状及未来趋势预测[J].山西农经,2018(12):127-128.
- [7] 李圆圆.基于 BP 神经网络的北京市二手房交易价格预测研究[D].首都经济贸易大学,2018.
- [8] 袁乖宁.数据挖掘技术在二手房市场评估中的应用研究[J].电脑迷,2018(05):232-233.
- [9] 米炜.探究二手房房价的影响因素[D].华中师范大学,2018.
- [10] 刘冰,金跃强,王书营.南京市二手房房价影响因素的多元线性回归分析[J].南京工业职业技术学院学报,2017,17(01):17-19.
- [11] 贾德铮,张思阳.基于房屋特征模型的房价影响因素研究[J].企业技术开发,2016,35(04):78-80+96.

附录：本案例所使用的 R 软件程序命令（部分）

第五节

`library(readxl)` % 读取 readxl 数据库

`data<-read_excel("C:/Users/Administrator/Documents/data1.xlsx")` % 读取本地的 Excel 数据文件

`attach(data)` % 将数据框添加到 R 的搜索路径 Excel 文件中

`hist(price, cex.main=1,col="light green",xlab="房价（元/平方米）",ylab="频数")` % 画直方图，并设定颜色，坐标轴的说明

`m<-mean(price)` ;m % 求房价的均值

`me<-median(price)`;me % 求房价的中位数

`boxplot(price~region,cex.main=1,col="light green",xlab="行政区域",ylab="区域平均价格（元/每平方米）")` % 邯郸市各行政区域所对应的房价箱线图

`hist(years,col="light green",cex.main=1,xlab="楼龄（年）",ylab="频数")` % 楼龄分布直方图

`boxplot(price~type,cex.main=1,col="light green",xlab="户型",ylab="房价（元/每平方米）")` % 各类户型所对应房价的箱线图

`boxplot(price~direction,cex.main=1,col="light green",xlab="朝向",ylab="房价（元/每平方米）")` % 各类朝向所对应房价的箱线图

`boxplot(price~decoration,cex.main=1,col="light green",xlab="装修程度",ylab="房价（元/每平方米）")` % 装修程度所对应房价的箱线图

`boxplot(price~propertyrights,cex.main=1,col="light green",xlab="产权性质",ylab="房价（元/每平方米）")` % 产权性质所对应房价的箱线图

第六节

`rm(list=ls())` % 清除 R 工作环境中的全部东西

`library(car)` % 加载 car 包

`library(readxl)` % 加载 readxl 包

`data1=read.data<-read_excel("C:/Users/Administrator/Documents/data1.xlsx")` % 读取本地的 Excel 数据文件

`attach(data1)` % 将数据框添加到 R 的搜索路径 Excel 文件中

(一)不考虑交互作用的多因素协方差分析模型

```
lm1=lm(totalprice~years+ area+ floor+ as.factor(apartment)+  
decoration+as.factor(region)+as.factor(direction)+as.factor(propertyrights))% 建立回  
归方程
```

```
anova(lm1) %输出关于回归方程的标准方差分析表
```

(二) 不考虑交互作用的多因素协方差修正模型

```
lm2=lm(totalprice ~ years+ area+floor+ as.factor(apartment)+  
decoration+as.factor(region)) %建立回归方程
```

```
anova(lm2) %输出关于回归方程的标准方差分析表
```

```
summary(lm2) %提取回归分析结果
```

(三) 考虑交互作用的方差分析

```
lm3=lm(totalprice~years+ area+ floor+ as.factor(apartment)+  
decoration+as.factor(region)+as.factor(direction)+as.factor(propertyrights)+  
as.factor(apartment)*as.factor(direction)+as.factor(propertyrights)*as.factor(region))  
%建立回归方程
```

```
anova(lm3) %输出关于回归方程的标准方差分析表
```

(四) 考虑交互作用的多因素协方差分析修正模型

```
lm4=lm(totalprice~years+ area+ floor+ as.factor(apartment)+  
decoration+as.factor(region) +as.factor(propertyrights)*as.factor(region)) %建立回归  
方程
```

```
anova(lm4) %输出关于回归方程的标准方差分析表
```

```
summary(lm4)
```

第七节

```
library(readxl) %引用数据库 readxl
```

```
data<-read_excel("C:/Users/Administrator/Documents/data.xlsx")% 读取本地的  
Excel 数据文件
```

```
attach(data)%将数据框添加到 R 的搜索路径 Excel 文件中
```

```
library(grpreg) %引用数据库 grpreg
```

```
library(ncvreg) %引用数据库 ncvreg
```

```
X=as.matrix(data[,c(1:6)]) %把 data 中的数据转换为矩阵，并赋值给 x
y=data$totalprice%把 data 中的 totalprice 赋值给 y
group<-c(1:6) %把 1 到 6 赋值给 group
lcol=c('yellow1','black','red',"violetred","tan3","steelblue2") %给 lcol 赋值 6 种颜色
fit<-grpreg(X,y,group,penalty="grLasso")%对 X, y 做 Lasso 回归并赋值给 fit
labs=c("1-years","2-area","3-floor","4-apartment","5-decoration","6-region") %提前
进行标签的设置
plot(fit,legend.loc="topleft",lwd=3,alpha=0.0002,legend=labs,col=lcol) %对 Lasso
回归方程画图，标签位置设置在左上角，线条宽度为 3，置信水平为 99.98%，数
据标签为 labs，颜色为 lcol。
cvfit<-cv.ncvreg(X,y,penalty="lasso") %交叉验证找 lasso 中最优的 lambda 值:
fit<-cvfit$fit%把 cvfit 中的 fit 赋值给 fit
beta<-fit$beta[,cvfit$ min] %选择 fit 中的最小 beta
```