### 摘 要

伴随着西部大开发，国家对于中西部地区的关注度持续上升，西安市作为十三朝古都，也是中西部地区的发展速度较快的城市之一，自然受到人们的青睐，更多人口迁入西安，加快西安的经济发展的同时，也加速了西安房价的抬升。房子在中国人心目中的地位不言而喻,房价快速上涨问题已经逐渐成为一个关系国计民生的热门话题。城市房价的过快增长,不仅直接影响了我国城镇居民的生活水平,还给社会经济发展带来了许多不稳定的因素。因此,准确预测我国城市房价已经逐渐成为国内学者研究的一个热点问题，本文便对与此热门问题相关的三个任务展开讨论与探究。

对于任务一，本文利用近几个月西安市内各小区的位置等信息以及对应价格和西安各区二手房价格的变化数据将影响西安二手房价格的因素进行了横向分析和纵向分析，建立了预测二手房房价的模型，解释了影响二手房价格的因素，为有意购买二手房的家庭提供了合理、有效的建议。

对于任务二，本文将基于任务一纵向分析得到的模型预测的西安各区五月份房价均值与西安五月份房价数据进行了对比和误差计算，也将基于任务一横向分析得到的模型预测的西安某些小区的房价与这些小区五月的实际房价进行了对比和误差计算。在此基础上建立了可以同时估测某月某小区的房价的改进后的新模型，并给出了应用实例。

对于任务三，本文建立了模型描述了政策改变影响后的房价变化趋势，并根据此模型预测了西安市二手房接下来一年的价格均值与成交量。

在文章的最后，本文针对前文所建立的模型进行了敏感度分析，并客观地分析了本文所建立模型的优缺点，指出了其实际应用价值。

### 问题重述

据华商报报道，5月西安新建商品住宅销售价格环比上涨0.3%，二手住宅销售价格环比下降0.4%。其中，新房价格涨幅相比上月有小幅提高。今年前5个月里，西安1月新房价格没有变化，2、3月分别上涨1%、0.4%，4月份涨幅回落到0.2%，5月上涨0.3%。今年5月西安新房市场整体呈现平销局面。而在5月28日晚，西安出台政策调整商品住房交易，涉及降低限购门槛、缩短限售时间等。随后，市面上有大量新房房源开始入场。预计6月西安新房供应量有可能达到上半年供应新高。值得注意的是，西安二手房价格走势，继续与新房价格走势出现背离。继4月环比下降0.1%后，5月份西安二手房降幅有所扩大，环比下降0.4%。分产品来看，5月份西安90㎡及以下、90-144㎡及144㎡以上二手房价格均环比下降，环比降幅分别为0.1%、0.5%和0.5%。西安二手房数据中心最新统计数据显示，市民选购二手房考虑的重点因素是地理位置，户型方面首选三居室。从购房者关注户型、面积区间及价格等方面看，市民浏览关注的主要户型是三居室，5居室以上的户型受关注度低。二手房地理位置也备受关注。选购二手房时往往希望离地铁近、附近有超市和学校、近市区。如果购房不是用来自住，地理位置也决定了房产的升值空间。

面对西安二手房市场众多的房源，购买者有自己的挑选标准。那么，影响二手房价格的因素有哪些，你又了解多少呢？

任务一：请根据（https://zjj.xa.gov.cn/esfgp/index.aspx）西安在售二手房和附件1的相关数据，建立数学模型，解释影响西安二手房价格的因素，为有意购买二手房的家庭提供建议。

任务二：请各队伍在网上找到西安二手房5月份的数据，检验任务一所获模型的合理性，并做改进。

任务三：西安市政府2022年5月28日出台“西安关于调整商品住房交易政策有关问题的通知”（参见附件1），该通知政策将对二手房价有怎样的影响？试分析并预测接下来的一年内西安二手房价和成交量的走势。

### 问题分析

题目向我们指出了影响西安二手房价的因素的部分因素和近年来西安市二手房、新房的房价涨势，本文的关键在于通过分析已有数据，建立数学模型来预测二手房价格变化，推断影响二手房价格的因素并根据该模型给出合适的购房建议。该模型由以下三部分组成：

2.1 对同一时期影响西安二手房价的因素进行横向分析

针对西安的房价评估，本文是使用的方法是建立多元线性回归模型，该回归模型以小区环境和教育质量、公共设施便利程度、户型为指标来西安市房价预测分析，因此商品房价格作为因变量，环境因素和户型因素作为自变量。本文的数据来源是2022年XJTU校内赛命题中B题的附件一和西安市住房和城乡建设局。

2.2 对不同时期西安二手房价的涨势的纵向分析

根据商品价格与商品价值的变化曲线，我们假设商品的价格围绕着商品价值周期性波动，使用傅里叶变换拟合数据，但同时，我们注意到房产的价值还会随着当地经济的发展随时间有线性变化，所以将傅里叶变换中的常数项改为一次函数，来表达这种趋势。对于个别地区的房价波动性不强，相反其线性增长的趋势明显，在剔除一个过于偏离的数据后，其线性增长的趋势更加显著，我们采用了灰色模型拟合数据。

2.3 对受政策因素影响后的西安二手房价变化趋势的分析

政策的改变会对房价带来阶跃式的变化，变化后的房价在一段时间后会趋于稳定，故我们假设政策对房产市场的影响为指数式衰减，在实施后的头一个月对房产价格有突跃影响，而在之后政策红利逐渐消失，房价再次回归其价值曲线。

### 问题假设

|  |
| --- |
|  |

1. 假设影响公共设施便利程度指标的因素是与地铁站和公园的距离，二者评分方式以x（x为该小区与地铁站或公园距离）为自变量的指数函数且按照1：1加权。
2. 假设商品价格围绕着商品价值周期性波动。
3. 假设政策给房价带来的影响指数式衰减。

|  |
| --- |
|  |

### 四、符号说明

|  |  |  |
| --- | --- | --- |
| **序号** | **符号** | **意义** |
| 1 | **p** | 房价预测值 |
| 2 | **x** | 月份 |
| 3 | **a0** | 线性增长系数 |
| 4 | **a1** | 房价波动系数 |
| 5 | **a2** | 房价波动系数 |
| 6 |  | 小区推荐房价 |
| 7 |  | 小区所在行政区域平均房价 |
| 8 |  | 小区推荐房价与小区所在行政区域平均房价的比值 |
| 9 |  | 房价 |
| 10 |  | 多元线性回归模型中的常数项 |
| 11 |  | 多元线性回归方程中的第一个系数 |
| 12 |  | 多元线性回归方程中的第二个系数 |
| 13 |  | 多元线性回归方程中的第三个系数 |
| 14 | **t** | 两个因子之间的相关性系数 |
| 15 | **Ses** | 多元线性回归方程的系数矩阵 |
| 16 | **STATS** | 多元线性回归模型的分析矩阵 |
| 17 | **BINT** | 系数估计值的置信边界下限和置信边界上限 |
| 18 | **R** | 残差 |
| 19 | **RINT** | 用于诊断离群值的区间 |
| 20 | **a** | 小区附近公共设施建设程度 |
| 21 | Der | 小区环境质量及配套服务程度 |

|  |
| --- |
|  |

### 模型的建立、求解与检验

5.1 任务一

5.1.1 户型对二手房价格的影响与提出建议

针对户型对西安二手房价的影响，本文是使用的方法是建立多元线性回归模型，来西安市房价预测分析，因此商品房价格作为因变量，环境因素和户型因素作为自变量。本文的数据来源是2022年XJTU校内赛命题中B题的附件一和西安市住房和城乡建设局，截取的是2022年7月份的相关数据。

根据题中“西安二手房数据中心最新统计数据显示，市民选购二手房考虑的重点因素是地理位置，户型方面首选三居室”内容得知，影响二手房价格的客观因素主要有地理位置和户型，通过调研，最受关注的户型是三居室、其次是二居室，一居室、四居室、五居室及其以上的户型关注度较少，且房源较少，通过统计计算，在二手房市场上，二居室的比例为0.34，三居室的比例为0.46，四居室的比例为0.09，一居室的比例为0.10。比例图如图所示。



图5.1.1户型比例图

对于地理环境的评估，为了适配多元线性规划模型，采用以下公式对各个小区的平均房价进行处理。

其中为此小区的参考价，为此小区所在区域的二手房均价，通过小区均价除以小区所在行政区域的二手房均价，可以对各个小区的地理环境因素进行一个量化评估，从而适配多元线性回归模型。当然，在这个地理环境因素中，有非常多其他的的影响因素，例如距离地铁站的远近距离、学区房的影响、购房政策的影响、采光的好坏、物业的服务建设程度和周边基础设施建设完备程度等等一系列因素，建立一个简便合理的模型是非常重要的，故忽略了每个因素单独的影响，将这些影响因素集成到的计算中，从而简化模型。

针对户型的量化评估，采用的是户型所占比例作为其特征值，从而载入到指定向量内进行matlab中多元线性回归模型的求解，这种用比例作为特征值的方法是较为简便和可行的。

通过以上两个特征值向量和具体到某个小区某个户型的价格向量，即一个小区一个户型对应的价格和其环境因素还有户型因素建立多元线性回归模型。

多元线性回归模型通过最小平方函数对多个自变量和因变量进行建模分析的一种线性回归模型，由于西安市房价是受地理环境和户型两个因素共同影响的，房价就是多元线性回归模型的因变量，两个影响因素作为自变量，建立的回归模型方程为：

其中，作为一个小区的特定户型对应价格的向量，即因变量，作为常数项，作为回归系数，含义是当X或变化一个单位时，价格的值会上升或下降或，t作为的相关系数，本文初步定t为1。

通过matlab，对上述模型进行求解，得到以下结果：

对结果进行分析，得到以下结果:

Ses=

3348.59490988616

10992.2845086212

-1303.45928765296

-95.4621690593402

最终，可以得到=3348.59490988616，10992.2845086212，-1303.45928765296和-95.4621690593402。分别对应以上多元线性回归的方程的各个系数。故最终多元线性回归方程为：

对结果画图得到：

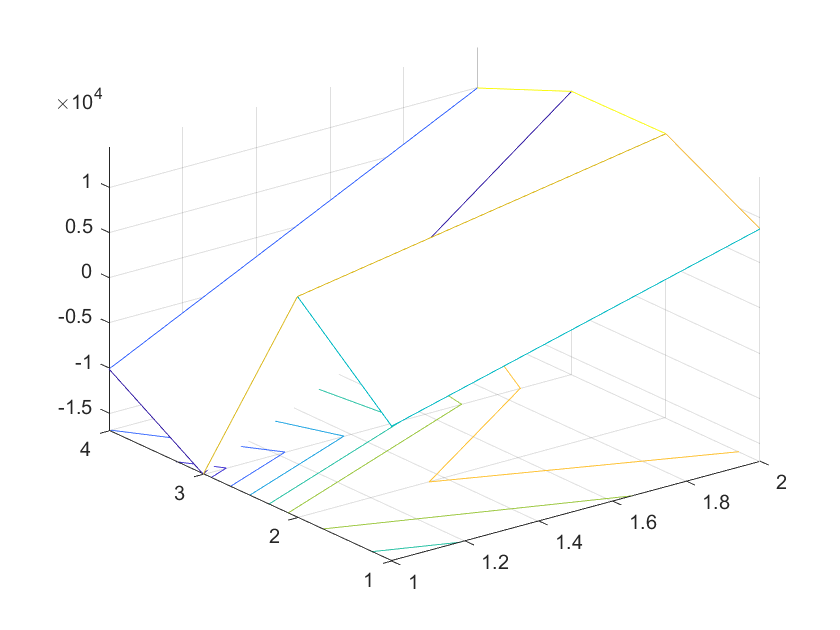


图5.1.2 bint系数估计值的上下限结果图

图5.1.3 R残差结果图

图5.1.4 离群值区间图

STATS =

0.841335996409592 70.7016905637911 4.84139180344693e-16 5946507.12151111

对所得模型结果进行画图可得图1、图2和图3。

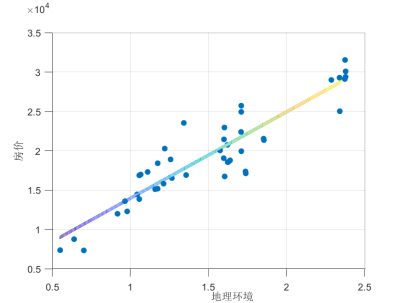
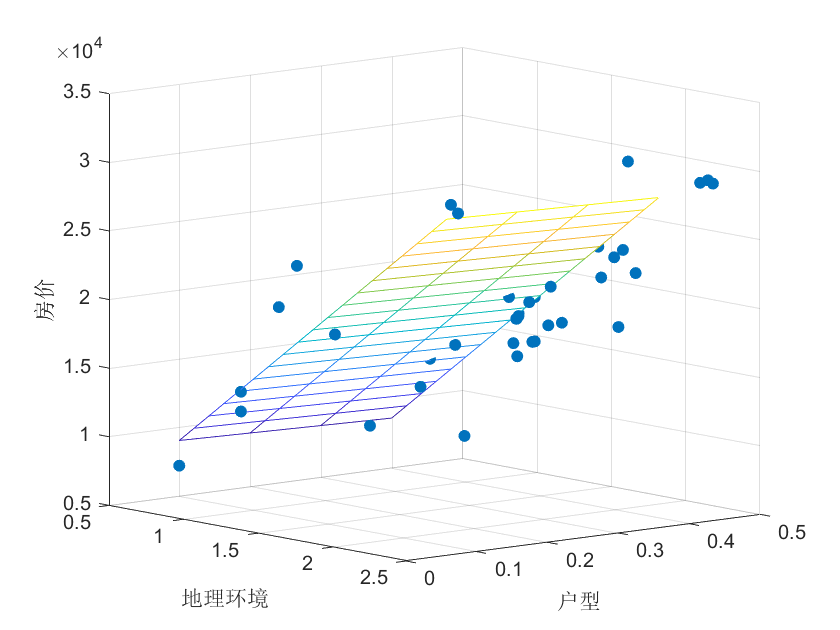
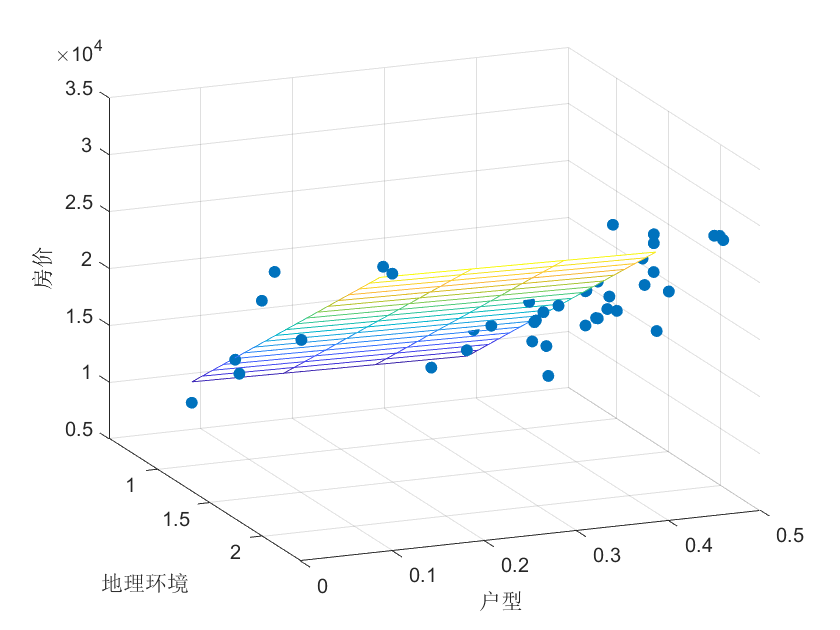


图5.1.5 模型结果图

与数据散点图（图4）进行比较，模型图更具有分析和比较价值。

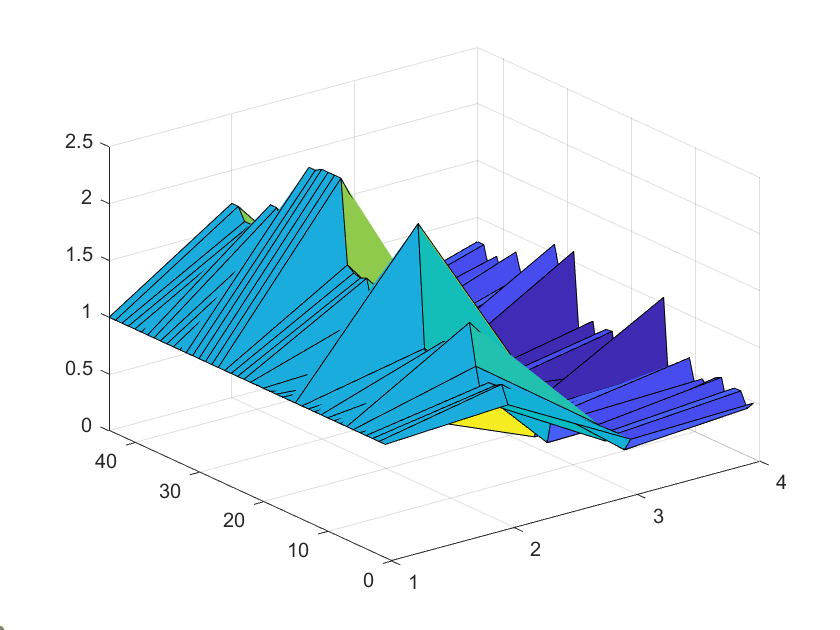


图5.1.6 数据散点图

以上结果，Ses为多元线性回归的系数估计值，以数值向量形式返回，BINT为系数估计值的置信边界下限和置信边界上限，以数值矩阵形式返回，R为残差，以数值向量形式返回，RINT为用于诊断离群值的区间，以数值矩阵形式返回，STATS模型统计量，以数值向量形式返回，包括统计量、F统计量及其p值，以及误差方差的估计值，ALPHA为显著性水平，这里定为0.05。

对模型结果进行分析，bint的第一列包含每个系数估计值的置信边界下限；第二列包含置信边界上限。如果X的列是线性相关的，则regress为bint中对应于b中零元素的元素返回零。



图5.1.7 系数估计值的上下限结果冷暖图

如果观测值i的rint(i,:)区间不包含零，则对应的残差大于100\*(1-alpha)%的新观测值的预期残差，即表明存在离群值。



图5.1.8 RINT矩阵的xbar图

对所得的结果进行分析，STATS向量里，0.841335996409592为，判定系数越接近于1，代表回归方程越显著，p值4.84139180344693e-16小于0.05的设置的显著性水平，因此响应y和X中的预测变量之间存在显著的线性回归关系，说明拟合程度较好。

通过建立多元线性回归模型对西安市的房价进行分析，可以得出随地理环境因素的增强，二手房的房价也随之增加，对于户型因素而言，户型对房价的总体影响不大，大多数人在选购二手房时都选择二居室或者三居室。

对于购买二手房的家庭，本文推荐首先将心仪的房子的每平方米价格、其所在小区的参考价和居室信息载入，进行matlab点画图，可以得出心仪房子在求解的多元线性回归模型中的位置，若在求解函数之下，则意味着心仪的房子物超所值，推荐购买；若在求解函数之上，则意味着心仪的房子卖的贵了，不推荐购买。通过这个分析方法从而综合考虑环境因素和户型因素，为有意购买二手房的家庭提供必要的建议。当然，模型仅考虑了环境因素变化和户型因素变化，时间和政策的变化集成在本文的傅里叶拟合模型中。

房地产行业是我国的支柱行业，房价增长速度过快会限制了人民的生活水平，也对我国经济的平稳发展起到了影响。从政府角度出发，应该合理规划建造安排，完善土地的开发制度。从居民角度出发，要结合实际，理性购房，不要盲目投资于房市之中，要小心谨慎对待房地产的价格变化。

5.1.2 小区公共设施便利程度与小区品质和教育质量对二手房房价的影响与提出建议

将本文附件一导入matlab，提取五个表格（table）类型的数据后，将所有的表类型转换为数值向量，才可以进入后续的计算。

这五个数值向量中，d1向量对应于距离小区最近地铁的距离，d2向量对应于距离小区最近公园的距离，der1对应于小区附近教育质量的评分，der2对应于小区本身质量的评分。

通过以下两个公式将d1和d2两个数值向量进行处理后。

再通过

将距离对小区房价的影响集成到一个向量里，从而使模型合理化。

对于der1和der2两个数值向量，直接通过

进行处理后，作为此小区房价的另一个影响因素。

至此，模型建立完成，采用的仍旧是多元线性回归模型，但是对于评估地理位置有了更加确切准确的系统，从而模型更趋合理化，考虑了更多的因素，让模型更加精准。

多元线性回归模型通过最小平方函数对多个自变量和因变量进行建模分析的一种线性回归模型，由于西安市房价的地理位置是受多个因素共同影响的，房价就是多元线性回归模型的因变量，多个影响因素作为自变量，建立的回归模型方程为：

其中，作为一个小区的对应价格的向量，即因变量，作为常数项，作为回归系数，含义是当a、der或变化一个单位时，价格的值会上升或下降。在这里，t仍作为1.

通过matlab，对上述模型进行求解，并采用画图的方式展示出来，得到以下结果：

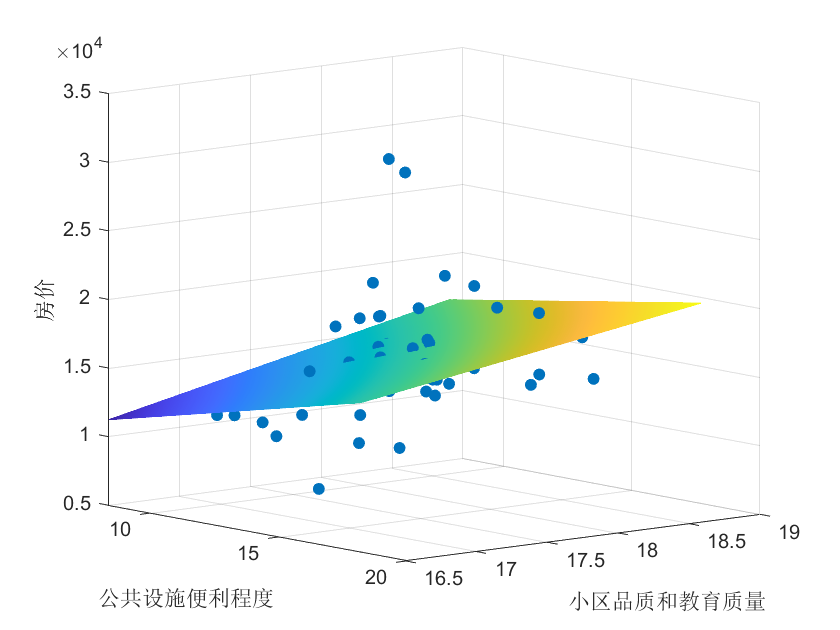


图5.1.9 新模型结果图

但是，这个模型中有一些散点过于离散，导致整个模型的相关性不高，其STATS向量为：0.0808014271759477 1.40646740871585 0.252366111695265 16468810.4835544，第一个0.0808014271759477为，判定系数越接近于1，代表回归方程越显著，可以看到未接近于一，导致其相关性不高，拟合效果不好。

对结果进行分析，是这些散点导致的问题，通过对RINT矩阵（RINT为用于诊断离群值的区间）画图，得到如下结果：



图5.1.10 RINT矩阵图

通过观察，其中第19、20、24、28、35、38和52点与拟合结果有较大偏差，其定义这几个点的数据为离群值后，再进行新一轮的模型优化。

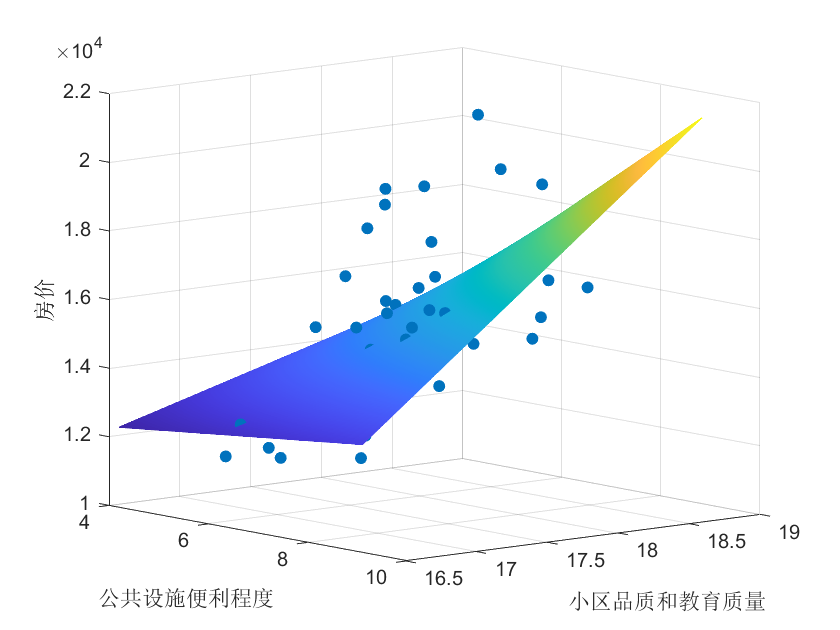


图5.1.11 删除部分散点后的多元线性回归模型图

图5.1.12 删除点后模型的RINT矩阵图

经过新一轮模型的带入，STATS矩阵为0.304746092855129 5.99042051575592 0.00175199471909096 4029052.82014359，其中明显相关性再提高。

通过观察，6、9、11、18、28、37和39点明显偏离，去除后，再次进行分析，得到以下结果。

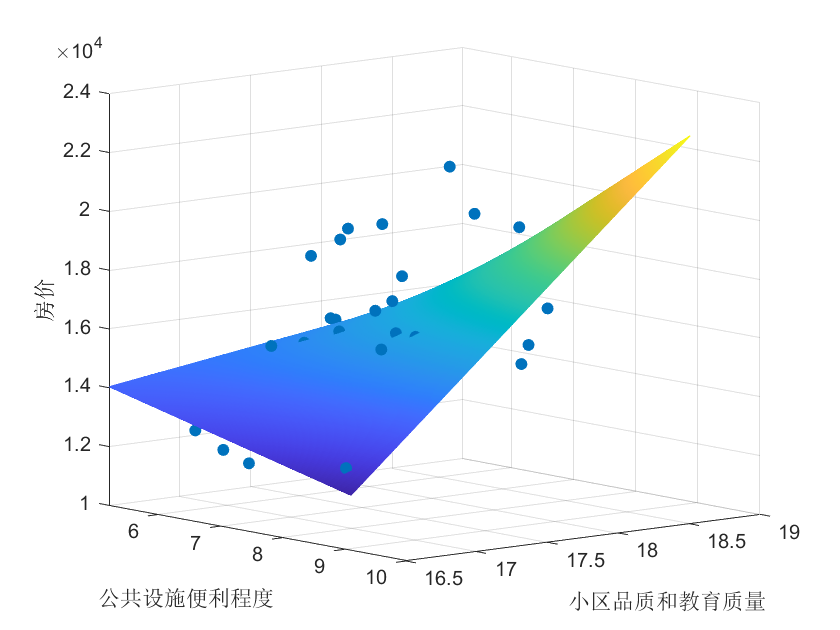


图5.1.13 优化后的多元线性回归模型

经过新一轮模型的带入，STATS矩阵为0.303290368184991 4.93360602820337 0.00596196361236421 4242596.39856170，其中明显相关性已不明显提高。

通过以上模型来看，线性回归效果不好的原因可能是数据量不足。通过补充数据量至76份，就已经出现了较好地相关性。其结果如下图所示。

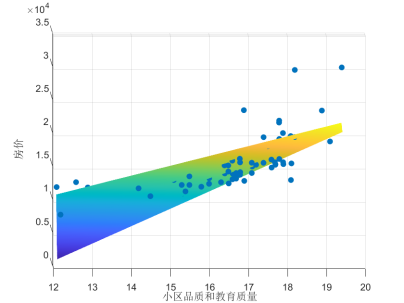
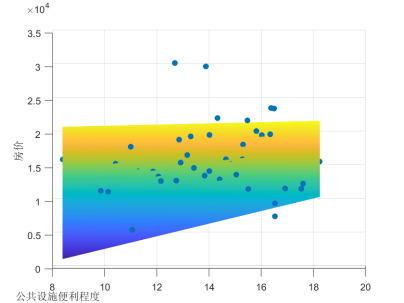
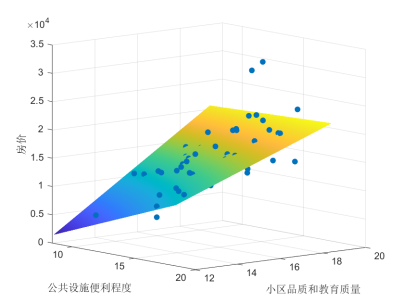


图5.1.14 补充数据量后的回归线性模型

再补充相关数据后，数据量为一百，得到了最终的数据图。

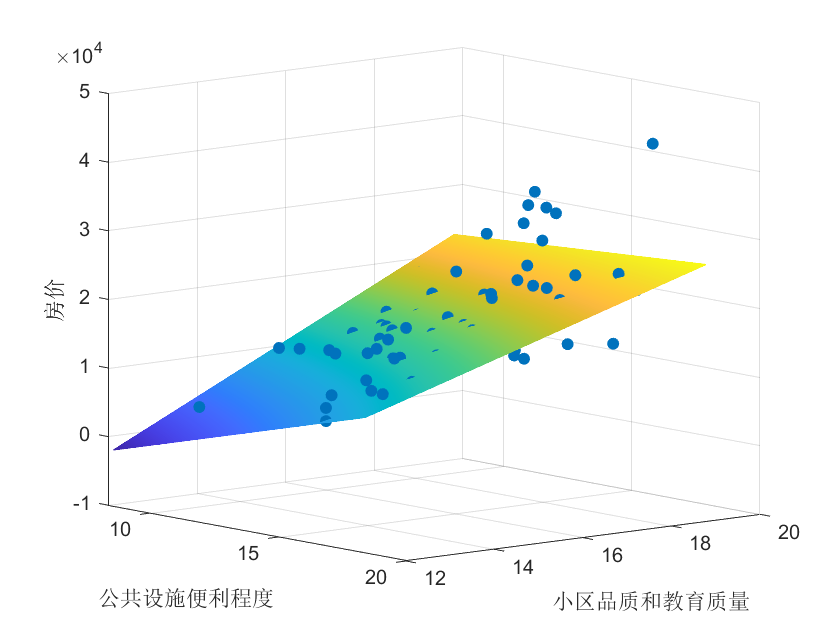


图5.1.15 数据量为100的回归线性模型

其STATS矩阵为0.556975136553649 40.2307090195067 6.32987980841137e-17 13876554.9344945，可以看到相关系数明显增大许多，进行了显著性分析，其6.32987980841137e-17也明显小于0.05。



图5.1.16 补充数据量后的RINT离群矩阵图

最终，得到的四个系数分别为-55766.2127807406，2224.01066127644，3790.64871404110，-95.3877492926356，得到的方程为：

r

经分析得出，产生这些离群的值小区大多在西安的“五大名校”附近，所以本文合理猜测小区房源是否是“学区房”的因素导致了小区的公共设施便利程度与环境舒适度等指标在一些如此类地特殊情况下并不能很大程度地影响小区的房价，所以二手房的购买者可以在没有对”学区房“的需求的条件下尽量在保证的公共设施便利程度与环境舒适度的前提下选择与”五大名校“较远的小区。

5.1.3 对西安各区二手房价格的预测模型与提出建议

根据商品价格与商品价值的变化曲线，我们假设商品的价格围绕这商品价值周期性波动，使用傅里叶变换拟合数据，但同时，我们注意到房产的价值还会随着当地经济的发展随时间有线性变化，所以将傅里叶变换中的常数项改为一次函数，来表达这种趋势。

在代码实现上，我们先通过二项傅里叶变化拟合得出房价波动频率和对应系数值，在进行线性拟合得到最终预测函数，以下其他各区预测函数获取方法同理。

未央区：

线性模型：

系数(置信边界为95%)

根据a0数据的置信边界，我们有超过95%的把握认为未央区房价以总体上升的趋势发展。加之，政府放松对购房的限制，会进一步促进房价上升，所以建议有在未央区购买房产需求的家庭尽早购买，或者等到政策再次加紧对房地产的调控时，购入最划算。

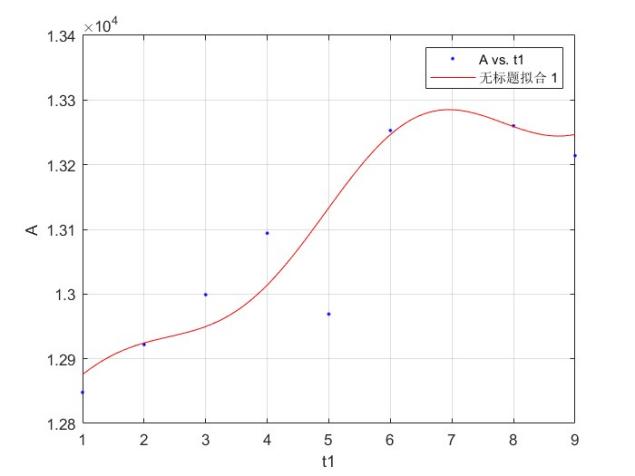


图5.1.17 未央区房价涨势图

雁塔区：

线性模型:

系数(置信边界为 95%):

根据a0数据的置信区间，我们有超过一半的把握认为雁塔区的房价会逐渐下降，但由于近期政策的放松，雁塔区的房价有反弹的趋势，所以建议在雁塔区有购房需求的家庭近期不要入手，等待政策重新收紧或者等待政策对房产市场促进作用消失后再入手。

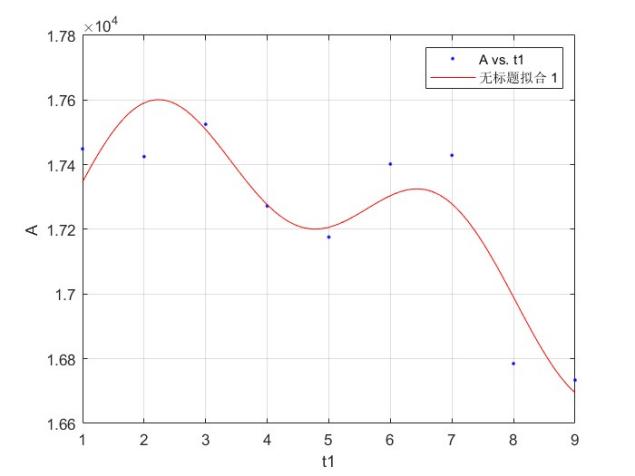


图5.1.18 雁塔区房价涨势图

莲湖区

线性模型:

系数(置信边界为 95%):

根据a0数据的置信区间，我们有超过一半的把握认为莲湖区的房价近期会逐渐上涨，加之新的宽松政策对房产市场的促进作用，本文建议最近有在莲湖区购入二手房需求的家庭可以在近期购入。

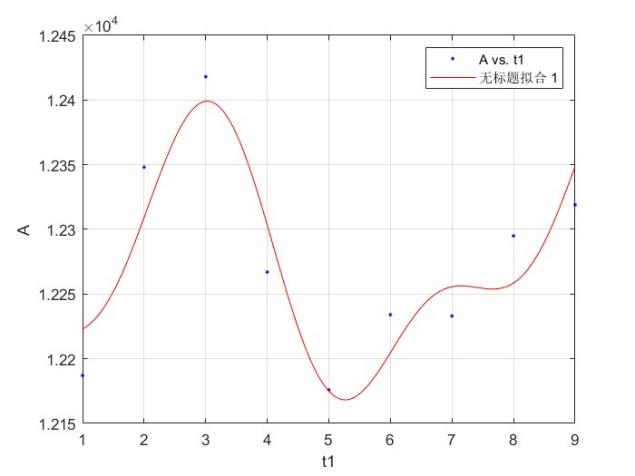


图5.1.19 莲湖区房价涨势图

新城区

线性模型:

系数(置信边界为 95%):

根据a0的置信区间，我们认为新城区房价未来一段时间内将保持温和上涨态势不变，建议在新城区有购房需求的在房价处于12000附近时可以考虑入手。

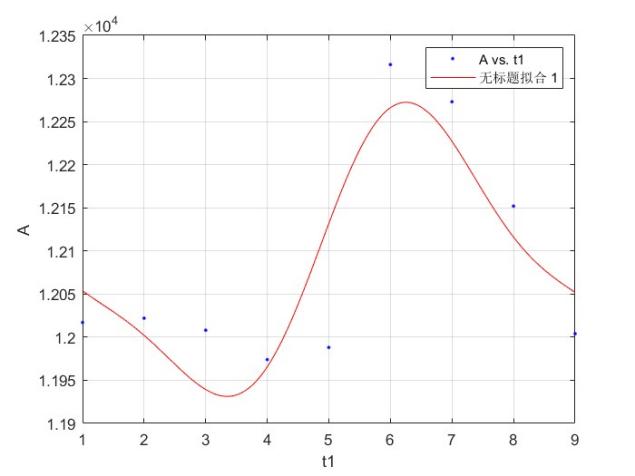


图5.1.20 新城区房价涨势图

碑林区

线性模型:

系数(置信边界为 95%):

根据a0数据的置信区间，且在宽松政策出台后，碑林区六月份房价也只温和上涨了0.38%，所以本文相信在最近一段时间内碑林区房价将继续维持低位，最近有在碑林区购房需求的家庭可以考虑入手。

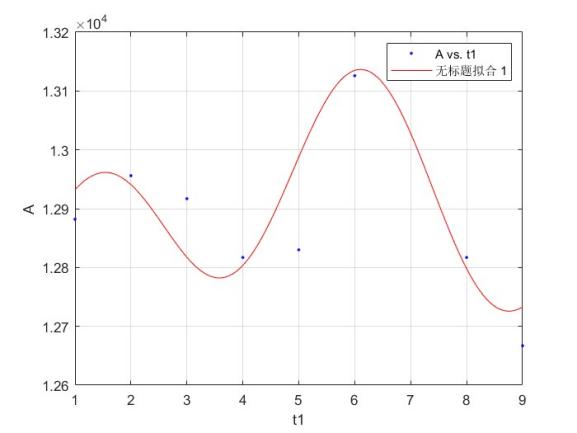


图5.1.21 碑林区房价涨势图

上述拟合方法对这四个区的房价有较好的拟合效果，但对于长安区和灞桥区拟合效果差强人意，原因是，最近一年来这两个区的房价变化数据波动性不强，相反其线性增长的趋势明显，在剔除一个过于偏离的数据后，其线性增长的趋势更加显著。而且，根据附件一所给数据，灞桥区二手房均价六月同比去年上涨7.56%，长安区二手房均价比去年同期上涨15.33%。均远高于新城区、碑林区、莲湖区、雁塔区和未央区，应采用另一种线性拟合模型。我们采用了灰色模型。

设时间序列：有n个初始值，通过累加生成新时间序列为，其中，，。

建立的紧邻均值生成序列为，其中，。

则GM(1，1)模型是由1个变量组成的一阶线性动态模型，其微分方程为：

 （1）

式中，是对原始数列的一次累加操作，a称为发展系数，b称为灰色作用量。

解微分方程（1）式，得到的时间响应式为：

 （2）

对应的时间相应序列为：

， （3）

引入参数向量，利用最小二乘法求出，计算公式为：

， （4）

式中，

， 

进一步求出（3）式的累减还原式为：

， （5）

计算出将的值代入（5）式，得到的预测值。

基于上述流程，进行MATLAB流程，则可对长安区、灞桥区房价进行预测，得到的灰色预测模型结果如下：

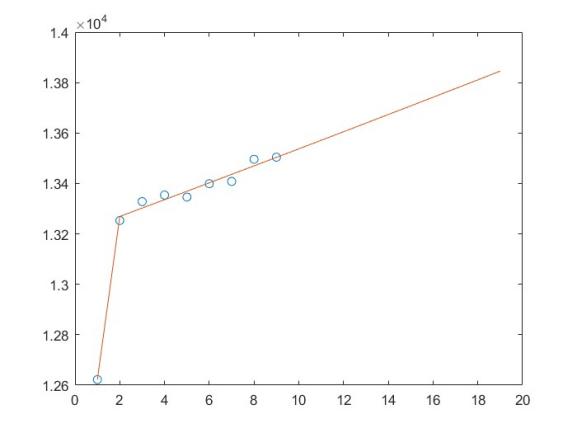
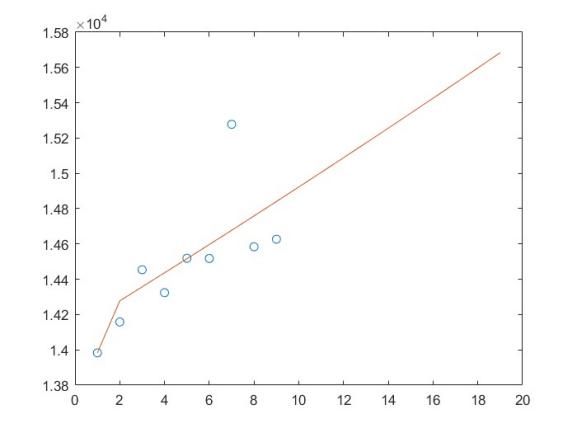


图5.1.22 长安区房价涨势图

图5.1.23 灞桥区房价涨势图

所以根据灰色模型的预测，这两个区的房价近期将持续保持高速上涨态势，而且在最近政策的支持下，房价不太可能出现下跌现象，所以本文建议有在灞桥区和长安区购房需求的家庭应尽快购入房产。

5.1.4 新房房价涨势图与二手房房价涨势图的对比分析

2021年8月到2022年7月西安二手房房价与新房房价走势图如下：

图5.1.24 二手房、新房涨势对比图

可以看出，二手房的房价走势图随着新房房价走势图的趋势变化，但波动性更大。新房市场发生由跌到涨、由涨到跌变化时，都会引起二手房对应的更大反应。但当新房保持价格温和上涨或者温和下跌时，二手房价格可能会发生周期性波动。

5.2 任务二

5.2.1 根据5.1建立的模型对西安市五月二手房房价的预测结果和误差

5.1.1与5.1.2对二手房房价的横向估测：

模型5.1.1：

得到=3348.59490988616，10992.2845086212，-1303.45928765296和-95.4621690593402。分别对应以上多元线性回归的方程的各个系数。故最终多元线性回归方程为：

模型5.2.2：

最终，得到的四个系数分别为-55766.2127807406，2224.01066127644，3790.64871404110，-95.3877492926356，得到的方程为：

对于大唐世家的一套三户型的房子，其推荐价格为15532元（来自附件一），第一个模型预测的结果为16599元，误差系数为6.43%；第二个模型预测结果为16713元，误差系数为7.07%。

对于华府西城港湾一区，其推荐价格为13588元（来自附件一），第一个模型预测的结果为13978元，误差系数为2.80%；第二个模型预测结果为13476元，误差系数为0.82%。

对于西市家园，其推荐价格为15991元（来自附件一），第一个模型预测的结果为16448元，误差系数为2.78%；第二个模型预测结果为16969元，误差系数为3.07%。

5.1.3对二手房房价的纵向估测：

模型预测雁塔区5月房价为16730，而实际房价为16669，相对误差为0.37%，合理。

模型预测未央区5月房价为13303，实际房价为，相对误差，合理。

模型预测莲湖区5月房价为12434，实际房价为12143，相对误差为2.4%，在允许的合理误差范围内。

模型预测碑林区5月房价为12855，实际为12636，相对误差为1.7%，在允许的合理误差范围内。

模型预测新城区5月房价为11995，实际房价为12038，相对误差为0.36%，合理。

模型预测长安区5月房价为13537，实际房价为13412，相对误差0.93%，合理。

预测结果灞桥区5月房价为14923，实际房价为14820，相对误差0.70%，合理。

5.2.2 模型的改进

本文5.1部分建立的模型5.1.2与5.1.3是独立的两个模型，为了将两个模型串联到一起，并可以预测未来某月份某小区的房价，我们将模型5.1.2的y换成了y1:

其中y0是该小区所处区域（西安市某区或整个西安市）的楼房均价，y为当月某小区的房价。

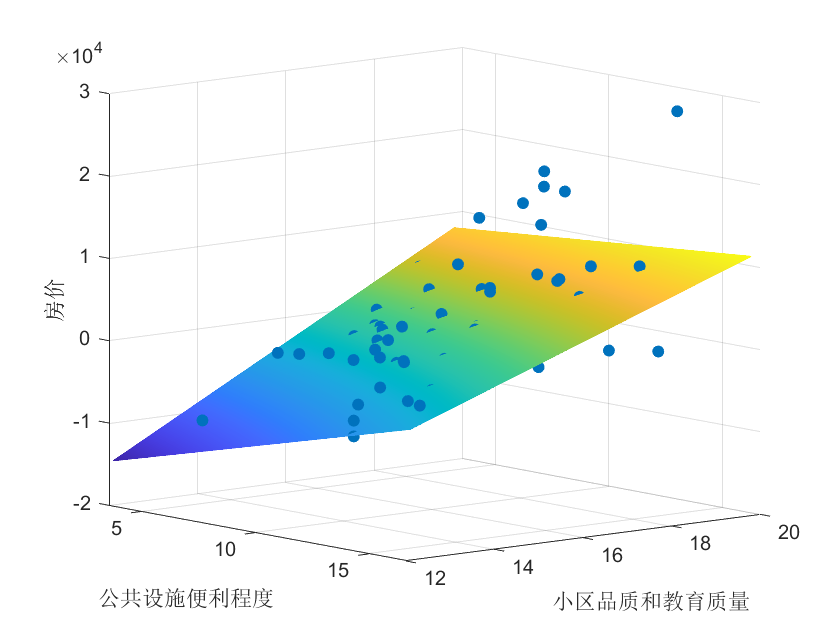


图5.2.1 改进后模型的结果图

基于改造后的模型对未来某月某小区的房价预测举例（预测2022年9月西市佳园小区的价格）：

利用5.1.3模型，得出莲湖区9月房价均值为12344元/平方米。

-56529.11037363131685.118778855023222.8702583110872.4428337649924

得出ρ=4330，算得西市佳园小区9月的价格为16674元/平方米。

5.3 任务三

我们假设政策对房产市场的影响为指数式衰减，在实施后的头一个月对房产价格有突跃影响，而在之后政策红利逐渐消失，房价再次回归其价值曲线。

对于未央区，根据上述模型，在没有政策改变影响的情况下，六月房价应为12839，但实际为13232，差值即为政策所带来的影响，所以得出其指数系数应为1068.3。

线性模型:

系数(置信边界为 95%):

根据价格和需求曲线，采用幂函数拟合的方法，可以获得房屋成交量与成交价格的反函数关系式，成交量

未央区未来一年的房价预测结果为：

表2 未央区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 13609 | 13545 | 13554 | 13644 | 13765 | 13834 |
| 27850 | 28840 | 28698 | 27325 | 25595 | 24665 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 13825 | 13798 | 13826 | 13914 | 13997 | 14036 |
| 24784 | 25145 | 24771 | 23633 | 22615 | 22153 |

对于雁塔区，根据上述模型，在没有政策改变影响的情况下，六月房价应为16547，但实际为16731，差值即为政策所带来的影响，所以得出其指数系数应为451.23。

线性模型:

系数(置信边界为 95%):

雁塔区未来一年的房价预测结果为：

表3 雁塔区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 16547 | 16383 | 16285 | 16316 | 16386 | 16324 |
| 6546 | 7048 | 7368 | 7265 | 7038 | 7238 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 16078 | 15781 | 15601 | 15555 | 15519 | 15411 |
| 8100 | 9300 | 10125 | 10349 | 10528 | 11087 |

对于莲湖区，根据上述模型，在没有政策改变影响的情况下，六月房价应为12331，但实际为12203，造成政策实施后实际销售量反而小于模型预测值的原因是，莲湖区五月房价出现突然跳水，而模型拟合没有考虑该数据，所以产生了误差我们将五月份数据带入，进行重新拟合，所得的函数R方小，拟合效果差。因此对于莲湖区房价我们采用灰色模型。

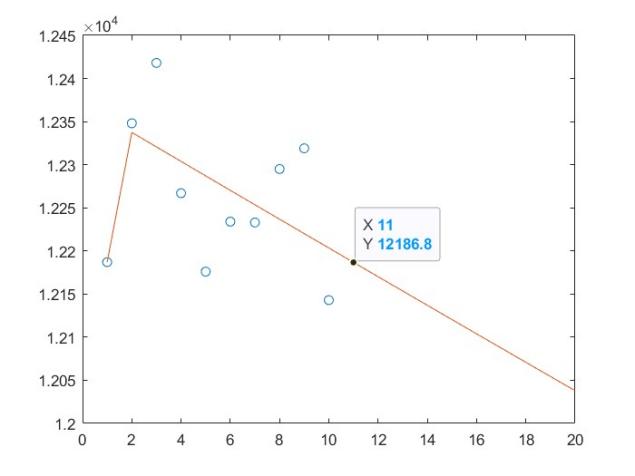


图5.3.1 莲湖区房价的灰色预测结果图

模型预测结果为12186，差值即为政策所带来的影响，所以得出其指数系数应为46.21。

表3 莲湖区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 12176 | 12155 | 12138 | 12120 | 12104 | 12087 |
| 63496 | 64313 | 64983 | 65702 | 66348 | 67042 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 12071 | 12054 | 12038 | 12022 | 12006 | 11990 |
| 67703 | 68413 | 69090 | 69774 | 70466 | 71165 |

对于新城区，根据上述模型，在没有政策改变影响的情况下，六月房价应为11945，但实际为12018，差值即为政策所带来的影响，所以得出其指数系数应为198.4。

线性模型:

系数(置信边界为 95%):

表4 新城区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 12049 | 12213 | 12289 | 12206 | 12104 | 12050 |
| 68624 | 62085 | 59297 | 62349 | 66348 | 68582 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 11987 | 11966 | 12091 | 12263 | 12287 | 12181 |
| 71297 | 72229 | 66878 | 60234 | 59368 | 63303 |

对于碑林区，根据上述模型，在没有政策改变影响的情况下，六月房价应为12920，但实际为12681，造成这种结果的原因和莲湖区一致，但这次将五月的意外数据带入后，所得到的拟合方程R方为0.94795，合理，可以接受该拟合函数。该拟合函数预测六月房价为12671，差值即为政策所带来的影响，所以得出其指数系数应为27.18。

线性模型:

系数(置信边界为 95%):

表5 碑林区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 12681 | 12557 | 12472 | 12580 | 12782 | 12821 |
| 46992 | 50540 | 53147 | 49859 | 44310 | 43321 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 12622 | 12385 | 12322 | 12387 | 12380 | 12252 |
| 48643 | 55976 | 58130 | 55909 | 56143 | 60636 |

对于灞桥区，应用灰色模型，可知对六月房价的预测为13515，而实际房价为13563，差值即为政策所带来的影响，所以得出其指数系数应为459.4。

对于长安区，应用灰色模型，可知对六月房价的预测为14959，而实际房价为15128，差值即为政策所带来的影响，所以得出其指数系数应为130.5。

表6 灞桥区未来一年的房价预测结果（表头分别为时间、房价、成交量）

表7 未央区未来一年的房价预测结果（表头分别为时间、房价、成交量）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2022年7月** | **2022年8月** | **2022年9月** | **2022年10月** | **2022年11月** | **2022年12月** |
| 13558 | 13572 | 13593 | 13618 | 13642 | 13668 |
| 28635 | 28417 | 28094 | 27714 | 27355 | 26972 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | | **2023年3月** | | **2023年4月** | | **2023年5月** | | **2023年6月** | |
| 13694 | 13720 | | 13746 | | 13772 | | 13798 | | 13824 | |
| 26595 | 26224 | | 25858 | | 25499 | | 25145 | | 24797 | |
| **2022年7月** | | **2022年8月** | | **2022年9月** | | **2022年10月** | | **2022年11月** | | **2022年12月** |
| 15097 | | 15134 | | 15196 | | 15267 | | 15343 | | 15410 |
| 12913 | | 12681 | | 12303 | | 11885 | | 11456 | | 11092 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **2023年1月** | **2023年2月** | **2023年3月** | **2023年4月** | **2023年5月** | **2023年6月** |
| 15498 | 15586 | 15674 | 15762 | 15850 | 15938 |
| 10634 | 10197 | 9781 | 9383 | 9004 | 8642 |

### 敏感度分析

本文对问题进行敏感度分析，综合考虑两个因子之间的相关性，即t系数。通过调节t系数，来确定t系数的变化对二手房房价的影响，即确定两个因子之间的相关性对模型求解结果的影响。



图6.1 第一个模型的敏感度分析



图6.2 第二个模型的敏感度分析

可以看出，当t增大时，第一个模型的每平方米房价也随之增加，增大到一定程度后不再增大；第二个模型刚好处于相反的趋势。同时，两模型的两个因子随t的变化也不尽相同。因此模型对于t的变化较敏感，若有能证明两个因子之间相关性会大大影响每平方米房价的证据，则t也应当是一个考察的因素。

由上述两个模型的图例可知，每次改变t值都可以稳定地得出对应的x1、x2和y的值，可见模型的稳定性较好。

### 模型的评价与推广

7.1 模型的优点

1.本文采用多个模型从不同维度来分析估算房价，条理清楚，符合实际情况，在预测与估算房价时误差较小。

2.本文结合客观实际，用较为贴近的模型来模拟了一些房价变换的趋势如政策影响下的房价变化趋势，所得到的结果符合实际大致趋势与大众的认知。

3.改进后的该模型的实用性较强，可以误差较小地预测未来某月某小区的房价、对小区房价设置的合理性进行评价、向购房者提供合理建议。

7.2 模型的缺点

1.由于对于数据预测的方法为直接拟合，该模型没有对新房房价涨势与二手房房价涨势关系的定量分析，无法将新房房价的波动纳入影响二手房房价的讨论范围内。

2.影响西安二手房房价的因素具有一定的特殊性（当地的学校受欢迎程度较为参差，导致教育方面对二手房房价的影响较大），而本文则未考虑此特殊性，导致在横向分析影响二手房房价的因素时采用的多元线性回归模型的相关程度不高。

7.3 模型的推广

1.模型可以推广到对于其他商品的价格预测与价格估算上。

2.模型可以推广到对于商品价格合理性的评价上，并可利用此模型算得某商品的“合理价格区间”。

### 参考文献

［１］杨中宣，杨洋洋．河南省房价影响因素及其预测研究［Ｊ］． 科技创新与生产力，２０１７（０７）：３３－３６＋３９．

［２］王树静．基于灰色理论的北京市商品住宅价格因素分 析及预测［Ｄ］．北京：北京理工大学，２０１５．

［３］王楠，郑晨曦．基于灰色系统 ＧＭ（１，１）的济南市商品房 房价预测模型［Ｊ］．中国管理信息化，２０１８，２１（１５）：１２９ －１３１．

［４］袁芳．基于 ＡＲＩＭＡ 模型的西安市新建住宅价格指数的 分析和预测［Ｊ］．中国科技信息，２０１３，１１：１８０＋１８２．

### 附件

附件1：5.1.1模型代码

c=0.09;

d=0.1;

price=[a;b;c;d];

x1=[1.0562;1.2103;1.1743;1.1738;0.9769;1.2651;1.6243;1.2552;1.0654;1.1081;1.3564;1.1580;1.0562;1.2142;2.2875;0.5439;0.9593;0.9117;0.7007;1.0410;0.6365;1.3364;1.5974;1.5726;1.5924;1.6365;2.3736;2.3736;2.3736;2.3736;2.3397;2.3397;1.7371;1.7371;1.7085;1.7085;1.8533;1.8533;1.6214;1.6214;1.6019;1.6019;1.7097;1.7097];

x2=[a;b;a;a;b;a;b;b;a;b;a;a;a;c;a;c;c;d;a;b;b;c;b;b;c;b;c;d;a;b;b;a;b;a;a;b;b;a;a;b;a;b;a;a];

y=[16875;15822;18417;15183;12295;16539;18570;18901;16988;17314;16917;15120;13863;20281;28997;7371;13597;11996;7340;14464;8768;23536;21451;20044;19055;18784;30086;29386;29117;31540;25027;29288;17127;17376;22398;19941;21364;21539;20751;18554;22955;16740;24944;25731];

X = [ones(size(x1)) x1 x2 x1.\*x2];

[ses,BINT,R,RINT,STATS] = regress(y,X,0.05)

scatter3(x1,x2,y,'filled')

hold on

x1fit = min(x1):0.01:max(x1);

x2fit = min(x2):0.01:max(x2);

[X1FIT,X2FIT] = meshgrid(x1fit,x2fit);

YFIT = ses(1) + ses(2)\*X1FIT + ses(3)\*X2FIT + ses(4)\*X1FIT.\*X2FIT;

mesh(X1FIT,X2FIT,YFIT)

xlabel('地理环境')

ylabel('户型')

zlabel('房价')

view(50,10)

hold off

附件2：5.1.2模型代码

d1= table2array(d1);

d2= table2array(d2);

y= table2array(y);

der1= table2array(der1);

der2= table2array(der2);

d1= table2array(d1);

d2= table2array(d2);

y= table2array(y);

der1= table2array(der1);

der2= table2array(der2);

a1=10.\*exp(-(d1./2000));

a2=10.\*exp(-(d2./2000));

a=a1+a2;

der=der1+der2;

X = [ones(size(a1)) a der a.\*der];

[ses,BINT,R,RINT,STATS] = regress(y,X,0.05)

scatter3(a,der,y,'filled')

hold on

afit = min(a):0.01:max(a);

derfit = min(der):0.01:max(der);

[AFIT,DERFIT] = meshgrid(afit,derfit);

YFIT = ses(1) + ses(2)\*AFIT + ses(3)\*DERFIT + ses(4)\*AFIT.\*DERFIT;

mesh(AFIT,DERFIT,YFIT)

xlabel('公共设施便利程度')

ylabel('小区品质和教育质量')

zlabel('房价')

view(50,10)

hold off

代码二：

a1=10.\*exp(-(d1./1000));

a2=10.\*exp(-(d2./1000));

a=a1+a2;

der=der1+der2;

y=y-14418;

X = [ones(size(a1)) a der a.\*der];

[ses,BINT,R,RINT,STATS] = regress(y,X,0.05)

scatter3(a,der,y,'filled')

hold on

afit = min(a):0.01:max(a);

derfit = min(der):0.01:max(der);

[AFIT,DERFIT] = meshgrid(afit,derfit);

YFIT = ses(1) + ses(2)\*AFIT + ses(3)\*DERFIT + ses(4)\*AFIT.\*DERFIT;

mesh(AFIT,DERFIT,YFIT)

xlabel('公共设施便利程度')

ylabel('小区品质和教育质量')

zlabel('房价')

view(50,10)

hold offs

-56529.1103736313

1685.11877885502

3222.87025831108

-72.4428337649924

附件3：5.1.3模型代码1（以未央区为例）

clear,clf,clc

syms a b;

c=[a b]';

A=[12848 12922 12999 13094 12969 13253 13376 13260 13214];

B=cumsum(A);

n=length(A);

for i=1:(n-1)

C(i)=(B(i)+B(i+1))/2;

end

D=A;D(1)=[];

D=D';

E=[-C;ones(1,n-1)];

c=inv(E\*E')\*E\*D;

c=c';

a=c(1);

b=c(2);

F=[];

F(1)=A(1);

for i=2:(n+10)

F(i)=(A(1)-b/a)/exp(a\*(i-1))+b/a;

end

G=[];G(1)=A(1);

for i=2:(n+10)

G(i)=F(i)-F(i-1);

end

t1=1:9;

t2=1:19;

G

plot(t1,A,'o',t2,G)

q=createFit(t1,A)

p=createFit6(t1,A)

function [fitresult, gof] = createFit2(t1, A)

%CREATEFIT2(T1,A)

% X 输入: t1

% Y 输出: A

% 输出:

% fitresult: 表示拟合的拟合对象。

% gof: 带有拟合优度信息的结构体。

[xData, yData] = prepareCurveData( t1, A );

% 设置 fittype 和选项。

ft = fittype( 'fourier2' );

opts = fitoptions( 'Method', 'NonlinearLeastSquares' );

opts.Display = 'Off';

opts.StartPoint = [0 0 0 0 0 0.628318530717959];

% 对数据进行模型拟合。

[fitresult, gof] = fit( xData, yData, ft, opts );

% 绘制数据拟合图。

figure( 'Name', '无标题拟合 1' );

h = plot( fitresult, xData, yData );

legend( h, 'A vs. t1', '无标题拟合 1', 'Location', 'NorthEast', 'Interpreter', 'none' );

% 为坐标区加标签

xlabel( 't1', 'Interpreter', 'none' );

ylabel( 'A', 'Interpreter', 'none' );

grid on

function [fitresult, gof] = createFit6(t1, A)

%CREATEFIT6(T1,A)

% X 输入: t1

% Y 输出: A

% 输出:

% fitresult: 表示拟合的拟合对象。

% gof: 带有拟合优度信息的结构体。

%% 拟合: '无标题拟合 1'。

[xData, yData] = prepareCurveData( t1, A );

% 设置 fittype 和选项。

ft = fittype( {'x', '(-99.58\*sin(0.605\*x)-186.6\*cos(0.605\*x))', '(-89.22\*sin(2\*0.605\*x)-7.258\*cos(2\*0.605\*x))', '1'}, 'independent', 'x', 'dependent', 'y', 'coefficients', {'a0', 'a1', 'a2', 'a'} );

% 对数据进行模型拟合。

[fitresult, gof] = fit( xData, yData, ft );

% 绘制数据拟合图。

figure( 'Name', '无标题拟合 1' );

h = plot( fitresult, xData, yData );

legend( h, 'A vs. t1', '无标题拟合 1', 'Location', 'NorthEast', 'Interpreter', 'none' );

% 为坐标区加标签

xlabel( 't1', 'Interpreter', 'none' );

ylabel( 'A', 'Interpreter', 'none' );

grid on

附件4：模型5.1.3代码2（以长安区为例）

clear,clf,clc

syms a b;

c=[a b]';

A=[13983 14158 14454 14324 14519 14518 15278 14584 14627];

B=cumsum(A);

n=length(A);

for i=1:(n-1)

C(i)=(B(i)+B(i+1))/2;

end

D=A;D(1)=[];

D=D';

E=[-C;ones(1,n-1)];

c=inv(E\*E')\*E\*D;

c=c';

a=c(1);

b=c(2);

F=[];

F(1)=A(1);

for i=2:(n+10)

F(i)=(A(1)-b/a)/exp(a\*(i-1))+b/a;

end

G=[];G(1)=A(1);

for i=2:(n+10)

G(i)=F(i)-F(i-1);

end

t1=1:9;

t2=1:19;

G

plot(t1,A,'o',t2,G)

附件5：改进后的5.1.2模型

yq=y-14188;

a1=10.\*exp(-(d1./1000));

a2=10.\*exp(-(d2./1000));

a=a1+a2;

der=der1+der2;

X = [ones(size(a1)) a der a.\*der];

[ses,BINT,R,RINT,STATS] = regress(yq,X,0.05)

scatter3(a,der,yq,'filled')

hold on

afit = min(a):0.01:max(a);

derfit = min(der):0.01:max(der);

[AFIT,DERFIT] = meshgrid(afit,derfit);

YFIT = ses(1) + ses(2)\*AFIT + ses(3)\*DERFIT + ses(4)\*AFIT.\*DERFIT;

mesh(AFIT,DERFIT,YFIT)

xlabel('公共设施便利程度')

ylabel('小区品质和教育质量')

zlabel('房价')

view(50,10)

hold off

代码6：敏感度分析

syms x1 x2 a %定义符号变量

f=-55766.2127807406+2224.01066127644\*x1+3790.64871404110\*x2-95.3877492926356\*x1\*x2\*a;

f=simplify(f) %化简目标函数

f1=diff(f,x1), f2=diff(f,x2) %求目标函数关于x1,x2的偏导数

[x10,x20]=solve(f1,f2) %求驻点

pretty(x10), pretty(x20) %以书写习惯的方式显示

subplot(121), fplot(x10,[0.002,0.02]), title('') %画x1关于a的曲线

xlabel('$t$','Interpreter','Latex')

ylabel('$x\_1$','Interpreter','Latex','Rotation',0)

subplot(122), fplot(x20,[0.002,0.02]), title('') %画x2关于a的曲线

xlabel('$t$','Interpreter','Latex')

ylabel('$x\_2$','Interpreter','Latex','Rotation',0)

dx1=diff(x10,t), dx10=subs(dx1,t,0.01), dx10=double(dx10)

sx1a=dx10\*0.01/4735

dx2=diff(x20,t), dx20=subs(dx2,t,0.01), dx20=double(dx20)

sx2a=dx20\*0.01/7043

F=subs(f,{x1,x2},{x10,x20}) %求关于a的目标函数

F=simplify(F)

figure, fplot(F,[0.002,0.02]), title('')

xlabel('$t$','Interpreter','Latex')

ylabel('$y$','Interpreter','Latex','Rotation',0)

Sya=-4735^2\*0.01/553641

f3=subs(f,{x1,x2,t},{4735,7043,0.011}); f3=double(f3) %计算近似最优利润

f4=subs(F,a,0.011); f4=double(f4) %计算最优利润

delta=(f4-f3)/f4 %计算利润的相对误差

数据来源：

[中指云-数据商城-房产交易 (cih-index.com)](https://mall.cih-index.com/house)

[未央房价网,2022未央房价走势图,西安未央二手房价格-安居客 (anjuke.com)](https://xa.anjuke.com/market/weiyangq/)

用于分析的数据：



