中南财经政法大学

本科课程论文

论文题目：基于广义线性模型的短租房房价影响因素研究

院系名称：统计与数学学院

专业名称：应用统计 1701

作者姓名：曹奕涵

作者学号：201721090024

2019~2020 第二学期

使用 LaTeX 撰写于 2020 年 6 月 8 日

摘　要

本文主要研究了美国短租房价格的主要影响因素，并基于广义线性模型的方法，合理解释了几大因素对短租房价格的影响方向，建立了合理的短租房价格指导模型。研究结果表明，AirBnB 短租房的价格主要受房屋地区、房屋大小以及房屋类型的影响较大。其中，受房屋类型的影响最大。这也说明，由于短租房并不是由企业等拥有，而是个人房主拥有并出租，所以形式可以多种多样，并不需要局限于酒店房屋等等。本文的研究结果表明，应当鼓励房屋的多样化结合，地区的相互平衡，才能更好地引导短租房市场的发展。

关键字:　广义线性模型　LASSO　短租房

目录

1 数据来源

本实验数据集来自 kaggle 上的主要城市 AirBnB 房价数据集。AirBnB 为一家主营旅行房屋短期租赁的社区公司，于 2008 年成立，目前已经称为大多数年轻人出游短租的首选，也成为了硅谷几大互联网企业中的一员。

自 2015 年 AirBnB 正式进入中国市场以来，旅行短租房行业在国内飞速发展。对于年轻人来说，短租房节省了预订酒店的时间，并且能够花更少的钱，获得更优秀的旅行环境，结识房东，更好地融入旅行目的地，感受当地人的生活氛围。这使得近年来，不论是使用短租房平台的游客，还是房东，人数都大大增加。很多人通过购买或者使用闲置的房产，对其进行重新装修，调整，使得原本老旧的房产“摇身一变”成为炙手可热的受欢迎的房源，进一步加强了城市的可持续发展性。

本文基于 AirBnB 在美国几大主要城市的短租房价格数据，使用广义线性模型对几个用户在选择房型时的主要影响因素进行分析，建立了合理的短租房房价定价模型，帮助消费者以及短租房主更好地了解国内的短租房市场，引导合理的定价。

1.1 数据介绍

本数据集从原数据集中筛选出了 3196 条数据作为分析样本。清理后的数据集共有 16 个变量，分别为对数处理后的房屋价格，房屋类型，最多可接纳的住户人数，卫生间个数，床类型，取消政策，清理费用，所在城市，房主的档案照片，房主是否有真实个人信息，是否可以非连续地预定，评论总数，平均评分，房间个数，床个数。除去对数处理的房屋价格后，该数据集的所有变量皆为名义变量。因为变量个数较多，所以本文将使用 lasso 广义线性模型对数据变量进行一定的筛选。

1.1.1 对数价格总体分析

对对数价格做分布图，查看对数价格的偏态状况：

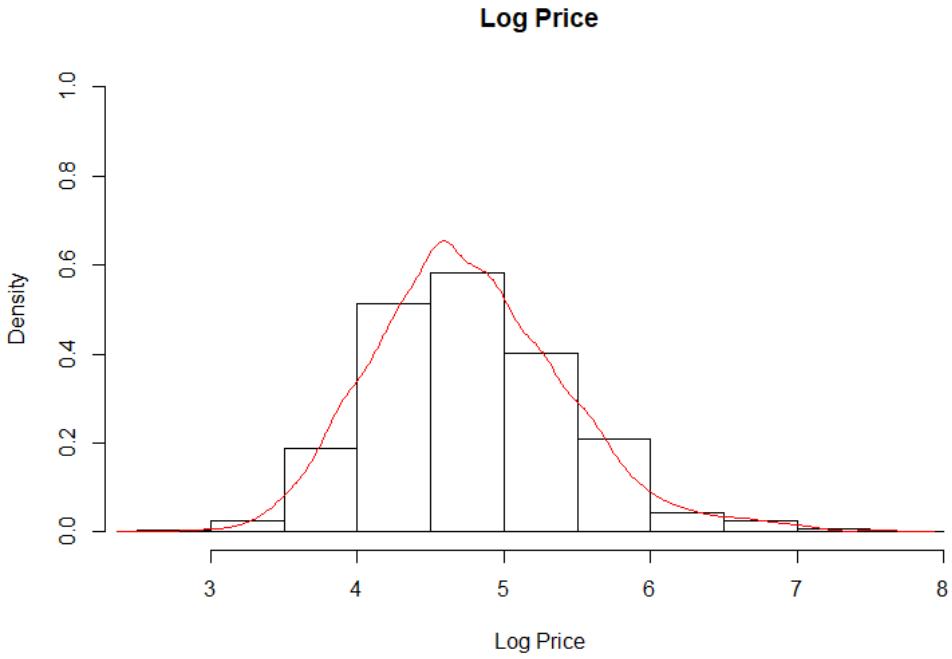

图 1.1: 对数价格分布图

可以看到，对数价格主要是呈右偏分布的，这也提示我们该数据集符合 gamma GLM 的分布情况。为了更清楚的分析该数据集，对其进行基础的描述性统计分析：

表 1.1: 对数价格描述性统计量

标准差	均值	中位数	最小值	最大值	偏度	峰度
0.67	4.77	4.7	2.71	7.59	0.53	0.66

可以看到，数据的确是呈右偏分布。最高的对数价格为 7.59，最低为 2.71。

1.1.2 地区价格分析

按预估，地区应该是一个对价格影响较大的因素，所以对数据按地区进行分类，查看各个地区的数据分布情况。对其进行分组的描述性统计，结果如下：

表 1.2: 对数价格描述性统计量

统计量/地区	数据个数	平均值	最大值	最小值	标准差	偏度
Boston	150	4.91	7.13	3.47	0.677	0.334
Chicago	155	4.65	7.09	3.22	0.608	0.479
DC	225	4.84	7.17	2.71	0.702	0.277
LA	956	4.72	7.58	3.22	0.670	0.668
NYC	1424	4.72	7.59	2.94	0.646	0.550
SF	286	5.16	7.09	3.81	0.615	0.557

根据简单的描述性统计分析的结果,本文中所涉及到的样本来自六个美国的主要城市——波士顿、芝加哥、华盛顿 DC、洛杉矶、纽约以及旧金山。其中，来自纽约的房屋占大多数，由于本文的数据是从原数据集中随意筛选的，这也从侧面表现出，短租房市场在纽约发展较好，市场较大，这与纽约作为国际大都市密不可分。6 个城市的短租房房价标准差差不多，说明几个城市房价的变动水平没有太大的区别。均值方面，可以看出，旧金山的短租房价格明显比其他城市要高一些。所有的城市的价格水平均呈右偏分布。对数据做直方图，进一步分析：

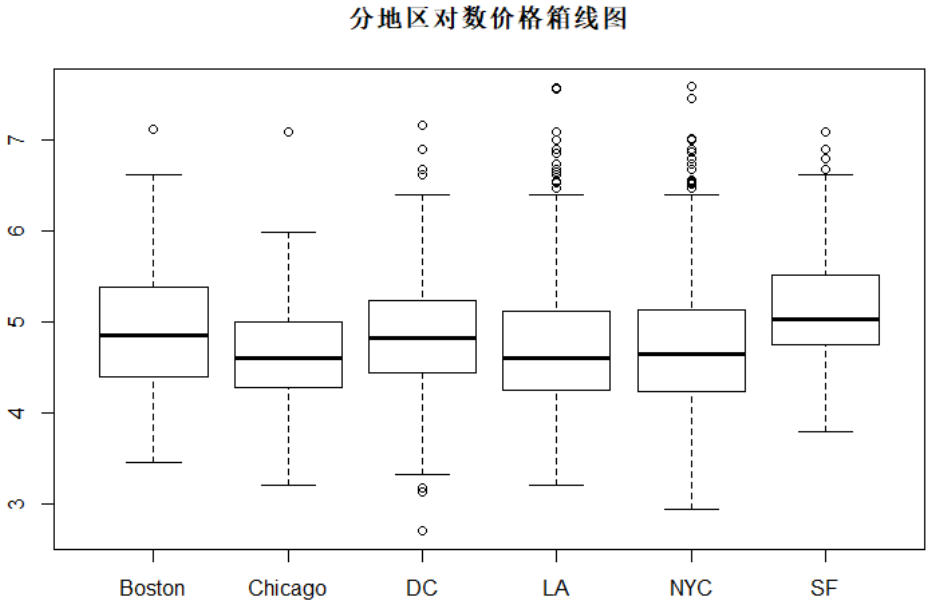

图 1.2: 分地区对数价格箱线图

根据箱线图，可以看出数据量较大的纽约与洛杉矶的离群点较多，而华盛顿 DC 的差值较大。排去离群点，房价最高值较大的是波士顿与旧金山，房价最低值较小的为纽约。总体来说，旧金山的整体房价水平还是偏上的。

1.1.3 可容纳人数分析

预估可容纳人数会对短租房的房价产生较大影响，故对可容纳人数进行分组，研究可容纳人数与对数价格的关系。

表 1.3: 可容纳人数分类的对数价格基本统计量表

可容纳人数	样本量	样本均值	样本最大值	样本最小值	标准差
1	323	4.102336855	6.549650742	2.708050201	0.455883591
2	1393	4.540381102	6.907755279	3.218875825	0.494018468
3	336	4.803902983	6.684611728	3.401197382	0.480125614
4	567	5.014960636	7.003065459	3.36729583	0.535685961
5	138	5.185139205	6.363028104	3.555348061	0.559793713
6	230	5.31431907	6.856461985	4.060443011	0.548193868
7	41	5.46262187	6.745236349	4.007333185	0.571743066
8	85	5.716334636	7.569411792	3.33220451	0.678126922
9	13	5.808694757	6.902742737	4.927253685	0.617220945
10	33	5.839671128	7.588323677	3.218875825	0.922152037
11	4	6.330481218	6.907755279	5.988961417	0.399031822
12	10	6.4144324	6.877296071	5.560681631	0.479024377
13	3	6.129759663	6.618738984	5.375278408	0.662885366
14	4	6.527053595	7.575584652	5.552959585	0.935754302
15	4	6.096706871	7.090076836	5.298317367	0.7607026
16	12	5.999488925	7.090076836	5.010635294	0.582321137

根据表1.4，绝大多数的短租房最多只能容纳 2 人。随着可容纳人数的上升，样本均值、样本最大值、样本最小值明显上升，说明短租房的房价的确会受房间可容纳人数的影响。

1.1.4 房间类型价格分析

预估房间类型会对短租房的对数价格产生较大影响，按房间类型（整个公寓/私人房间/共享的房间）分类进行讨论。分类型的对数价格分析结果如下：

表 1.4: 分房间类型的对数价格描述性统计量

房间类型	样本量	平均值	最大值	最小值	标准差
Entire home/apt	1870	5.12612199458663	7.588323677	3.583518938	0.568216693152681
Private room	1255	4.29831128470199	6.907755279	3.17805383	0.425830904182947
Shared room	71	3.85121300580282	5.010635294	2.708050201	0.49404264640093

绝大多数的房间都是按整套公寓的规格出租，只有小部分是共享类型的公寓。根据分类分析的结果，很明显，当房间类型为整套或者私人房间时，短租房的价格较高，而当房间类型为共享房间时，房间价格的样本均值、最大值、最小值都有明显的下降。

根据上文的分析，可以大致确定对房间价格有主要影响的几个因素为房间类型、可容纳人数以及房间的所在地区。

2 模型原理

2.1 Lasso 选择变量

Lasso 回归是回归估计的一种，是弹性网回归一种特殊情况，由斯坦福大学 Robert Tibshirani 教授于 2011 年在皇家统计上首先发表[1]。Lasso 在英文中的意思是“套索”，但作为统计方法，Lasso 代表的是 The Least Absolute Shrinkage and Selection Operator。Lasso 回归通过设定特殊的目标函数，在普通最小二乘估计的基础上根据这个新的目标函数估计参数，然后筛选出目标函数不为 0 的变量，这些变量就是对因变量有实际影响的。

那么，要如何设置目标函数从而对自变量进行筛选呢？在普通最小二乘法中，目标函数设定为：

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{2.1}$$

这样估计出来的参数就是最小二乘法的估计结果，也就是将目标函数设定为最小化 MSE，而 Lasso 的思想就是在其基础上，加上一个惩罚项，这个惩罚项能够约束参数，让不那么重要的变量的参数估计值很小，趋近于 0。所以，原目标函数变为：

$$\begin{aligned} \hat{\beta}_{lasso} &= \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 \sum_{j=1}^{p} x_{ij}\beta_j)^2 \\ subject \quad to \quad & \sum_{j=1}^{p} |\beta_j| \le t \end{aligned} \tag{2.2}$$

也可以表示为：

$$\hat{\beta}_{lasso} = \arg\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{2.3}$$

这两个公式表示的是同一个意思，就是在原来的普通最小二乘的目标函数上增加一个惩罚项，其中 λ 是一个很大的数，如果要最小化该式，就需要让 $\sum_{j=1}^{p} |\beta_j|$ 很小，那么就要最小化不重要的变量的参数值，就达到了筛选变量的目的。

2.2 广义线性模型

2.2.1 指数族分布

对于一个分布，如果它有如下形式：

$$p(y;\eta) = b(y)\exp(\eta^T T(y) - a(\eta)) \tag{2.4}$$

就称该分布属于指数分布族（exponential family）。

其中，参数如下：

- η 为自然参数
- $T(y)$ 为充分统计量（一般来说，$T(y) = y$）
- $a(\eta)$ 为对数配分函数（保证分布的形状）

这些定义，均与之后的广义线性模型的主要内容有关，可以说，GLM 就是基于指数分布族建立的。

2.2.2 GLM 模型

令 $X \in R^p$, Y 为观测值。假设这样一个线性模型：

$$E(Y|X) = \beta^T X \tag{2.5}$$

其中，$\beta \in R^p$。回忆逻辑回归，我们观测到的变量为某事件，发生或不发生，记作 $Y \in \{0, 1\}$，那么逻辑回归的模型则为：

$$\log(\frac{P(Y=1|X)}{1-P(Y=1|X)}) = \beta^T X \tag{2.6}$$

实际上我们对比 (2) 式和 (3) 式，不难发现，$P(Y=1|X) = E(Y|X)$，所以这里是在 (2) 式的基础上套了一个 $logit$ 函数。故广义线性模型也可以表达为：

$$g(E(Y|X)) = \beta^T X \tag{2.7}$$

这是一个对条件期望的变换。

下面，我们通过指数族分布，来了解一个广义线性模型的几大参数。

2.2.3 GLM 参数

GLM 模型的参数总共分为三个组成部分：随机构成（random component）：给定了 $Y|X$ 的分布；系统构成（systematic component）：将一个参数 η 与数据 X 联系起来的部分；以及一个连接函数（link function），将这两个部分组合起来的部分。这一概念，对应式2.7就非常好理解。

- 随机构成部分：具体化了一个条件分布。

  比如线性回归中，我们假设 $Y|X \sim N(\mu, \sigma^2)$，这就是一个随机构成项。在 GLM 中，我们一般假设 $Y|X$ 服从指数族分布：

  $$f(y;\theta,\phi) = \exp(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)) \tag{2.8}$$

  这里，我们把 θ 称作是自然参数，ϕ 为分散参数（dispersion parameter）。很好理解，现在考虑一个正态的情况，那么我们总是有 $\theta = \mu, \phi = \sigma$，这两个参数定义了 $Y|X$ 的形状。这里注意，θ 我们是不作为主要参数放到 GLM 模型中去的，有时候会看到是因为在一些模型，比如正态，$\mu = \theta$，但这并不一定是这样。

  一般来说，我们把 $E(Y|X)$ 记作一个参数 μ。也就是：

  $$E(Y|X) = \mu \tag{2.9}$$

  做 GLM 的目的就是去估计 μ，一般来说，ϕ 是已知的。

- 系统构成部分：将一个参数 η 与数据 X 联系起来。

  假设 η：

  $$\eta = \beta^T X = \beta_1 X_1 + \cdots + \beta_p X_p \tag{2.10}$$

  这就是一个系统性构成部分。

- 连接函数：

  连接函数将随机部分与系统部分连接起来：

  $$g(\mu) = \eta \tag{2.11}$$

  这就是一个广义线性模型。就是把2.7式中的几个组成用参数或函数的形式进行了替换。

2.2.4 Gamma GLM

本文中选择的广义线性模型为 log 连接的 Gamma 广义线性模型。也就是依照一下假定建立的广义线性模型：

$$\begin{cases} Y_i \sim Gamma \\ \log(\mu_i) = \log(E(Y_i)) = X_i^T \beta \end{cases} \tag{2.12}$$

3 实证分析

3.1 数据预处理

实验数据中存在着较多的问题，需要对数据进行进一步的清洗。首先，注意到数据中存在一些缺失值，但不会对数据有太大的影响，故本文中对存在有缺失值的数据进行了删除处理。删除缺失值数据之后，还剩下 3187 个样本。由于本数据的前后跨越时间并不长，所以忽略掉数据集中的最后评论时间这一变量。这样，最后待分析的数据变量共有 16 个，样本 3187 个。

本文中的数据大多为定性数据，需要对数据设置哑变量。设置完哑变量之后，原数据的变量个数拓展到了 50 个。这就更需要合理的选择变量的方法。

注意到，评分这一列的数据大多位于 80-100 之间，这样的分布可能会导致最后的估计出现问题，错误地删除这一变量。故需要对其进行标准化。

3.2 Lasso 选择变量

首先，使用 gaussian-Lasso 筛选对价格有较大影响的变量。按照 lasso 模型拟合得到的 lambda 值，按照最小的 lambda 值得出 50 个变量的估计系数。从中去掉所有为 0 的变量，结果如下：

表 3.1: Lasso 变量估计

变量	估计系数
(Intercept)	4.118
Boat	0.367
Boutique.hotel	0.115
Bungalow	-0.120
Camper.RV	-0.293
Castle	0.033
Condominium	0.101
Dorm	-0.037
Guest.suite	-0.128
Hostel	-0.263
House	-0.030
Loft	0.258
Other	0.033
Timeshare	1.061
Entire.home.apt	0.607
Shared.room	-0.399
Airbed	0.008
Futon	-0.114
strict	0.019
super_strict_60	1.476
Boston	0.009
Chicago	-0.240
DC	-0.051
LA	-0.117
SF	0.293
accommodates	0.066
bathrooms	0.178
data...8.	-0.008
hos1_has_pro0ile_pic	-0.234
hos1_iden1i1y_veri0ied	0.002
ins1an1_bookable	-0.034
review_scores_rating	0.042
bedrooms	0.123
beds	-0.010

对上面的系数估计结果进行整合，并去掉其中绝对值较小的变量，最后得到的最终变量共有 20 个，为：

表 3.2: Lasso 变量估计（筛选后）

变量	估计系数
super_strict_60	1.476
Timeshare	1.061
Entire.home.apt	0.607
Boat	0.367
SF	0.293
Loft	0.258
bathrooms	0.178
bedrooms	0.123
Boutique.hotel	0.115
Condominium	0.101
accommodates	0.066
Futon	-0.114
LA	-0.117
Bungalow	-0.12
Guest.suite	-0.128
host_has_profile_pic	-0.234
Chicago	-0.24
Hostel	-0.263
Camper.RV	-0.293
Shared.room	-0.399

这 20 个变量来自取消政策、住房类型、城市、浴室个数、卧室个数、可容纳人数、房主信息这几个名义变量。筛选下来的变量是符合事前分析的结果的，保留筛选下来的数据，对其进行进一步的分析。

3.3 模型拟合

用 Gamma GLM 对数据进行模型拟合分析，即 $Y_i \sim \Gamma(\mu_i, v)$，其中，连接函数为 $\log(\mu_i) = X_i\beta$。选取 Gamma 分布的理由为，数据值皆为正数，呈现较为明显的右偏分布。且在对数作用下，数据的方差也是接近常数的。并且，在这里对比 Gaussian 与 Gamma 两种分布的拟合情况，Gamma 分布的拟合结果 AIC 值比 Gaussian 分布小很多。

根据上文的分析，确定的广义线性模型如下：

$$\begin{aligned}\log(\mu) =&\beta_0 + \beta_1 cancellation + \beta_2 Timeshare + \beta_3 EntireHouse + \beta_4 Boat + \beta_5 SF + \beta_6 Loft \\ &+ \beta_7 bathrooms + \beta_8 bedrooms + \beta_9 BoutiqueHotel + \beta_{10} Condominium + \beta_{11} accommodates \\ &+ \beta_{12} Futon + \beta_{13} LA + \beta_{14} Bungalow + \beta_{15} GuestSuite + \beta_{16} Host\_profile \\ &+ \beta_{17} Chicago + \beta_{18} Hostel + \beta_{19} CamperRV + \beta_{20} Sharedroom\end{aligned} \tag{3.1}$$

软件对数据进行拟合的结果如下：

表 3.3: 参数估计结果

参数	估计值	标准差	t 值	p 值	
(Intercept)	1.441	0.036	39.902	2.1E-282	***
super_strict_60	0.264	0.088	2.991	2.8E-03	**
Timeshare	0.212	0.051	4.162	3.2E-05	***
Entirehomeapt	0.135	0.004	36.379	2.0E-242	***
Boat	0.093	0.039	2.360	1.8E-02	*
SF	0.063	0.006	11.195	1.5E-28	***
Loft	0.060	0.011	5.432	6.0E-08	***
bathrooms	0.032	0.004	8.388	7.4E-17	***
bedrooms	0.023	0.003	8.057	1.1E-15	***
Boutiquehotel	0.059	0.062	0.952	3.4E-01	
Condominium	0.029	0.009	3.301	9.8E-04	***
accommodates	0.012	0.001	10.230	3.5E-24	***
Futon	-0.033	0.017	-1.989	4.7E-02	*
LA	-0.026	0.004	-7.171	9.2E-13	***
Bungalow	-0.034	0.022	-1.571	1.2E-01	
Guestsuite	-0.040	0.031	-1.270	2.0E-01	
host_has_profile_pic	-0.064	0.036	-1.792	7.3E-02	.
Chicago	-0.052	0.007	-6.942	4.7E-12	***
Hostel	-0.072	0.044	-1.612	1.1E-01	
CamperRV	-0.080	0.039	-2.035	4.2E-02	*
Sharedroom	-0.104	0.011	-9.591	1.7E-21	***

4 分析结果

根据表3.3中的结果，假设给定 7% 的显著性水平，对短租房价格影响显著的因素有：

- 取消政策

  在 AirBnB 中可以选择房间的取消政策，也就是如果行程临时出现变化时，可以根据房主规定的房间取消政策进行调整。一般来说，取消政策较为严格的房屋是不允许免费取消的，常常需要索赔。根据分析的结果，系数为正数，这说明取消政策非常严格的时候，房价依旧会上升。这说明，对于整体水平较高的房屋，房价并不会因为取消政策的宽松而上升。

- 房间类型

  AirBnB 中可供选择的房间类型有很多，包括分时房、船屋、酒店式公寓等等。根据分析结果，这些房间类型的系数大多为正数，这说明，当房屋为分时房、船屋、酒店式公寓以及豪华酒店等等时，房屋的价格常常会增加，而当房屋为客房、旅馆、房车时，房屋的价格会减少。

- 地理位置

  总的来说，地理位置的影响并不是非常的大，这可能与样本量的大小有关。根据分析结果，当地点为旧金山时，估计值为正数，而当地点为洛杉矶、芝加哥时，估计值为负数。也就是说，旧金山的短租房价格普遍偏高，而洛杉矶、芝加哥的房价价格则普遍偏低。这与之前描述性统计分析的预估结果一样。

- 出租类型

  对于价格影响非常大的一个因素就是房屋的出租类型。在前面的描述性统计分析中就可以看到，当房屋为整租时，房价常常会较高，而当房屋为合租房时，房屋的价格则会降低很多。广义线性模型的分析结果与这一结果相符合。

- 房屋大小

  对于一个短租房而言，主要决定其价格的，还有一个必不可少的因素就是房屋的大小。本文主要通过可容纳人数、卫生间、房间个数来衡量一个房屋的大小。根据参数的估计结果，可以看到当这三者都增加时，房屋的价格也会增加，说明当房屋的大小越大时，房屋的价格常常更高。

根据本文的分析结果可知，对房屋价格影响最大的因素还是房屋的类型、大小。对于AirBnB 上的短租房而言，甚至用户评分、用户评论都不一定是最重要的，这可能是因为大多数的房屋并未收到很多的差评。当房屋的整体水平都较高时，这就要求房屋的种类等等能够“别出心裁”。

根据分析结果，对于国内的短租房市场，应当注重短租房的多样化发展，将短租房发展为一个能够体现地方“文化底蕴”的产业。并且，多地短租房产业应当相互扶持，相互促进，这样才能够促进短租房产业在国内的发展。

参考文献

[1] Robert Tibshirani. Regression shrinkage selection via the lasso. Journal of the Royal Statistical Society Series B, 73:273–282, 06 2011.

附录 A 附录

A.1 代码

```
# Set working dir

setwd('C:/Users/Administrator/Desktop/2020courses/final/dingxing')

# Read Dataset

#library(data.table)
#data <- fread('./data/new-york-city-airbnb-open-data/AB_NYC_2019.csv', sep

data <- read.csv('./data/airbnb.csv')
head(data)

hist(data$log_price, freq = FALSE, main = 'Log Price', xlab = 'Log Price',
lines(density(data$log_price), col = 'red')

library(psych)
describe(data$log_price)

#dest description
dest <- data.frame(data$log_price, data$city)
table(dest$data.city)
library(dplyr)
dest %>%
  group_by(data.city)%>%
  summarize(count = n(), mean = mean(data.log_price), max = max(data.log_pr

boxplot(dest$data.log_price ~ dest$data.city, main = '分地区对数价格箱线图'

#accommodates description
accom <- data.frame(data$log_price, data$accommodates)
res1 = accom %>%
```

```
  group_by(data.accommodates)%>%
  summarize(count = n(), mean = mean(data.log_price), max = max(data.log_pr

#room type
room <- data.frame(data$log_price, data$room_type)
res1 = room %>%
  group_by(data.room_type)%>%
  summarize(count = n(), mean = mean(data.log_price), max = max(data.log_pr

res1
write.csv(res1, 'res1.csv')

#glm

#glm(log_price~., family = Gamma, data = data)

library(glmnet)
library(nnet)
?glmnet
dummyV1 <- class.ind(data$property_type)
dummyV2 <- class.ind(data$room_type)
dummyV3 <- class.ind(data$bed_type)
dummyV4 <- class.ind(data$cancellation_policy)
dummyV5 <- class.ind(data$city)
glmnet(data[,2:16], data[,1], family = 'gaussian')
data[, 14] <- scale(data[,14])
data2 <- data.frame(data[,1], dummyV1 , dummyV2, dummyV3, dummyV4, dummyV5,
data3 <- na.omit(data2)
data4 <- na.omit(data)
fit1 = glmnet(as.matrix(data3[,2:50]),as.matrix(data3[,1]), family = 'gauss
fit2 = cv.glmnet(as.matrix(data3[,2:50]),as.matrix(data3[,1]), family = 'ga
coe <- coef(fit1, s = fit2$lambda.min)
summary(fit1)
plot(fit1)
fit1
act_index = which(coe!=0)
```

```
act_index
coe <- as.matrix(coe)
write.csv(coe, 'coef2.csv')

data5 <- data.frame(data3[,1], data3$super_strict_60, data3$Timeshare, data
names(data5)[names(data5) == 'data3...1.'] <- 'log_price'
# model
gammafit = glm(log_price~.,data = data5,  family = Gamma(link = 'log'))
lmfit1 = glm(log_price~., data = data5, family = gaussian)
res1 = predict(lmfit1,data5[,2:21])
plot(res1, lmfit1$residuals)

res2 = summary(gammafit)
write.csv(res2$coefficients, 'coef3.csv')

x <- abs(rnorm(100, mean = 0, sd = 3))
res <- abs(rnorm(100, mean = 0 , sd = 1))
y <- 1/x+res
fake <- data.frame(x,y)
glm(y~x, data = fake, family = Gamma(link = log))
```