

二手房市场观察与数据分析报告

陈冠华

二手房是指在房地产交易中心备过案，完成初始登记和总登记的，而后再次上市进行交易的房产。对二手房交易行业的研究之所以重要，在理论上是因为二手房可以促进地产行业发展，增加供给市场的多样性：产业提供了更多商品供给，丰富了居民的购房选择；二手房的价格往往低于新建住宅，因此也能帮助满足各种收入群体的置业需求。一个成熟的二手房行业能使住房市场形成阶梯消费模式，扮演好一、二、三级市场之间承前启后的角色。而在现实中，近些年密集的政策调控导致行业投资收窄，地产新开工增速逐步放缓，自2016年起房地产行业开始进入“高原期”。根据数据平台“移动观象台”的测算，2016年北京二手房交易量达到了新建住宅的2.3倍。在目前，二手房贡献了大部分的居住地产交易，而在未来这一比例还会进一步增加。因此二手房日趋成为行业焦点，研究二手房对判断中国房地产的现状与发展趋势具有很强的参考价值。

图一 北京市2016年二手房成交量分布图

二手住宅成交量 = 2.3倍 新建住宅成交量



¹2016年北京市二手房交易量达到了新建住宅的2.3倍

决定二手房价格的首先是其作为新建房源时的交易价格，而影响后者的因素可被分为经济因素（土地供给，人口数量与密度，GDP，建材成本等），区域因素（所在城市经济规模，是否属于学区房，周围配有的基础设施是否完善）与政策因素（政府是否限制了交易价格与交易量）等。此外，由于二手房的交易是一个信息不对称的过程，中介也会成为交易中的关键变量之一。市场的信息不对称程度越高，二手房成交价与其第一次成交时价格的偏差就越大，同时交易费用（中介费）也同成交价成正比。在狗熊会提供的北京市二手房交易数据集

¹ 引自《2016北京二手房客群深度洞察》<http://mi.talkingdata.com/report-detail.html?id=474>

中，CATE, subway, school 变量均属于区域因素，AREA 属于经济因素，该数据可以帮助我们更好地理解区位因素对房价的影响。

对 price 变量做描述性统计，得到以下结果：

表一 price 的描述性统计指标表

指标	最小值	下四分位数	中位数	均值	上四分位数	最大值	标准差	离散系数	极差
值 (元)	18348	42812	57473	61152	76100	149871	22293.3	0.36	131523

Price 的均值为 61152 元，中位数是 57473 元。均值高于中位数，结果说明 price 的分布呈右偏，部分高价房屋拉高了均值。使用描述性统计指标时，若极端情况较少，均值可以客观地反应数据的平均情况，而在极端情况比较严重时中位数更客观。

再对城区做分组统计，与图二对比可以发现北京市城区里的二手房交易量主要与城区面积和区位因素有关。石景山面积小，与核心功能区并不接壤，因此交易量最小；西城区与东城区都是核心功能区，因此虽然面积小但成交量仍然巨大。丰台，朝阳，海淀都属于城市功能拓展区，再加上面积大，因此排名比较靠前。

表二 北京市各城区二手房交易量表

城区	样本量
石景山	1947
西城	2750
东城	2783
朝阳	2864
海淀	2919
丰台	2947

图二 北京市城区图



平均房价如下所示，影响着城区均价的显然依旧是区位因素。越接近市中心的城区在城市规划中扮演的角色往往越重要，资源越多，基础设施也越完善，因此价值越高。在表格中作为首都核心功能区的东城区与西城区单价最高，海淀作为众多名校所在地以学区房的优势排名第三。

表三 北京市各城区二手房交易均价

城区	均价（元/平方米）
石景山	40286.89
丰台	42500.90
朝阳	52800.62
海淀	68757.60
东城	71883.60
西城	85674.78

教育资源一直是购房者最关注的因素之一。房屋周边有学校的比率如下所示，将表四中城区的顺序与表三做对比可以发现除了海淀上升了一位之外其他城区的排名完全相同。这也从数据上印证了教育资源与房价之间的强正相关性。

表四 北京市各城区二手房周边存在教育资源比率

城区	比率（%）
石景山	0.92
丰台	3.19
朝阳	20.84

东城	45.81
海淀	47.48
西城	56.11

房屋基础配套设施是否完善也很重要。对配套设施中最有代表性的地铁设施做分析，将表五中城区的排名与表三对比，发现两者基本一致。离市中心越近房屋周边有地铁的比率就越高，这是由城市建设侧重于核心区域所导致的。房屋周边有地铁设施代表交通发达，能够更好地节约购房者的出行成本，因此价值也更高。该指标和房价也呈比较强的正相关性，不过与表四比较可以看到，房屋拥有教育资源的比率远低于地铁，这说明了目前教育附加值在房地产市场中的稀缺性。

表五 北京市各城区二手房周边存在地铁设施比率

城区	比率 (%)
石景山	63.12
丰台	75.60
海淀	79.86
朝阳	88.65
西城	89.13
东城	94.90

附录：

R 语言代码如下所示：

```
dat = read.csv(
  "D:/meeting/在线实习项目/数据可视化与报告撰写/TASK2-统计图规范与常用统计指标/"
  header = T,
  stringsAsFactors = F
)
head(dat)
summary(dat$price)
boxplot(dat$price)
str(dat)

library(plyr)
region.n = ddply(dat, .(CATE), nrow)
colnames(region.n)=c("城区", "样本量")
region.n[order(region.n$样本量),]

region.price=ddply(dat,.(CATE),function(x){mean(x$price)})
colnames(region.price)=c("城区","房价")
region.price[order(region.price$房价),]

region.school=ddply(dat,.(CATE),function(x){mean(x$school)})
colnames(region.school)=c("城区","学校比率")
region.school[order(region.school$学校比率),]

region.sub=ddply(dat,.(CATE),function(x){mean(x$subway)})
colnames(region.sub)=c("城区","地铁比率")
region.sub[order(region.sub$地铁比率),]
```