

## 合租模式下的租房价格分析

**数据介绍：**data.csv，本案例的数据来源于某租房平台，共采集了北京市某年某月 5149 条合租房源的信息。本案例针对合租房间进行分析，若同一套房中有多个待租的房间，这些房间在本案例的数据中会对应多条数据，每一条数据对应其中一个合租房间，并且这些房间的数据中房源整体的信息相同（如房屋结构、地理位置等），但租赁面积、月租金不同。具体数据说明表如下：

| 变量类型 |      | 变量名        |        | 详细说明        | 取值范围                              |
|------|------|------------|--------|-------------|-----------------------------------|
| 因变量  |      | rent       | 季均销量   | 定量变量，单位：元   | 1150~6460                         |
| 自变量  | 内部结构 | area       | 租赁房间面积 | 定量变量，单位：平方米 | 5~30                              |
|      |      | room       | 租赁房间类型 | 定性变量，2 个水平  | 主卧、次卧                             |
|      |      | bedroom    | 卧室数    | 定量变量，单位：个   | 2~5                               |
|      |      | livingroom | 厅数     | 定量变量，单位：个   | 1~2                               |
|      |      | bathroom   | 卫生间数   | 定量变量，单位：个   | 1~2                               |
|      |      | heating    | 供暖方式   | 定性变量，2 个水平  | 集中供暖、自采暖                          |
|      |      |            |        |             |                                   |
|      | 外部条件 | floor_grp  | 所在楼层   | 定性变量，3 个水平  | 高楼层、中楼层、低楼层                       |
|      |      | subway     | 邻近地铁   | 定性变量，2 个水平  | 是、否                               |
|      |      | region     | 所在城区   | 定性变量，11 水平  | 朝阳、海淀、东城、西城、昌平、大兴、通州、石景山、丰台、顺义、房山 |

### 分析任务：

1. 请完成数据读入与汇总统计；
2. 绘制总体数据中月租金（rent）的直方图，观察月租金的大致分布，并进行简要解读；
3. 计算不同城区(region)租房平均价格，选取平均租金最高的 8 个城区，绘制降序平均租金柱状图，并作简要分析；
4. 绘制月租金（rent）-城区（region）分组箱线图，分析不同城区的房价差异，并给出简要解读。
5. 建立以月租金（rent）为因变量，其余为自变量的线性回归模型，来分析合租租房月租金的影响因素，观察最终的回归系数并尝试对系数进行解释。  
（提示：需要提前为因子类型指定基准组。

- \* 租赁房间变量以“次卧”作为基准组
- \* 所在楼层分组变量以“低楼层”作为基准组
- \* 邻近地铁变量以“否”作为基准组
- \* 城区变量以“石景山”作为基准组
- \* 供暖方式变量以“自采暖”作为基准组)

6. 对上面的回归模型利用 **BIC** 准则进行变量选择并解读结果。

7. 对 **BIC** 选择后的模型进行五折交叉验证，评估模型结果，并进行解读。

提交方式：以 HTML 的格式提交，建议使用 **Rmarkdown** 输出结果。注意：代码注释详细、简洁；解读详细、合理、得当。