

金融岗位招聘分析

数据介绍：本数据来自某大型招聘网站 2017 年 1 月在北京地区的金融行业招聘信息。数据包含 16195 条互联网招型聘信息,主要的特征如下：

1. 公司名称 “company”；
2. 公司属性 “nature”：国企、民营、股份制企业、上市公司、合资、外商独资、其他；
3. 公司规模 “size”：10000 人以上、1000-999 人、500-999 人、100-499 人、20-99 人、20 人以下；
4. 公司行业大类 “sector”：基金/证券/期货/投资、保险、信托/担保/拍卖/典当、银行；
5. 招聘岗位 “position”：具体的招聘岗位名称；
6. 岗位职责 “descrip”：岗位职责描述；
7. 薪资 “salary”：岗位平均薪资；
8. 学历 “education”：不限，大专，本科，硕士及以上；
9. 经验 “experience”：不限，1-3 年，3-5 年，5-10 年，10 年以上。

分析任务：

1. 请完成数据读入与数据概览
2. 对薪资(salary)做对数变换，并记为新变量(lnsalary)
3. 绘制总体数据中对数薪资(lnsalary)的直方图，观察薪资的大致分布，并进行简要解读
4. 绘制不同公司属性(nature)下，对数薪资(lnsalary)箱线图，并进行简要解读
5. 用所有的数据，将对数薪资(lnsalary)作为因变量，其余变量（除：公司名称、招聘岗位、岗位职责）作为自变量做线性回归来分析薪资的影响因素，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。
 - 公司属性变量以“国企”作为基准组
 - 公司规模变量以“20 人以下”作为基准组
 - 公司行业大类变量以“银行”作为基准组
 - 学历变量以“本科”作为基准组
 - 经验变量以“1-3 年”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 选择后的模型用所有的数据进行五折交叉验证，评估模型结果，并进行解读。