



Contents n t U[-]

- ▼ 5.2 成组样本比较的非参数检验
 - 5.2.1 两样本比较
 - 5.2.2 多样本比较
- ▼ 5.3 配对/配伍样本比较的非参数检验
 - 5.3.1 配对样本
 - 5.3.2 配伍样本
- 5.4 秩变换分析
- 5.5 实战练习
- ▼ 6 卡方检验
 - 6.1 卡方检验的基本原理
 - ▼ 6.2 行*列表的卡方检验
 - 6.2.1 scipy的实现方式
 - 6.2.2 statsmodels的实现
 - 6.2.3 事后两两比较
 - 6.3 卡方校正与确切概率法
 - ▼ 6.4 配对卡方检验
 - 6.4.1 基本原理
 - 6.4.2 配对卡方检验的统计量
 - 6.5 分层卡方检验
 - ▼ 6.6 二项分布检验与近似正态分布
 - 6.6.1 率的可信区间
 - 6.6.2 二项分布检验
 - 6.6.3 近似Z检验
 - 6.7 实战练习

6.7 实战练习

计算北京、上海、广州三地的汽车拥有率可信区间。

考察不同收入级别的受访者其职业分布有无差异。提示：需要考虑两两比较。

在上面分析的基础上，在控制城市的影响之后，考察不同收入级别的受访者其职业分布有无差异。

7 相关分析

7.1 相关分析的指标体系

7.2 相关分析的实现

相关分析作为比较简单的方法，在statsmodels中并未作进一步的完善，因此主要使用scipy实现

两个连续变量，且符合双变量正态分布：Pearson相关系数

```
scipy.stats.pearsonr(a, b)
```

两个连续变量，不符合双变量正态分布：Spearman等级相关系数

```
scipy.stats.spearmanr(a, b)
```

两分类变量 vs. 连续变量：Point-biserial相关系数