



## Contents n t U[-1]

- 1.1 中国消费者信心指数项
- ▼ 1.2 连续变量的统计描述
  - 1.2.1 直接使用汇总函数
  - 1.2.2 describe命令
  - 1.2.3 statsmodels的实现
- ▼ 1.3 分类变量的统计描述
  - 1.3.1 单变量的频数统计
  - 1.3.2 交叉表
- 1.4 实战练习
- ▼ 2 均数间的比较
  - 2.1 假设检验的基本原理
  - ▼ 2.2 单样本t检验
    - 2.2.1 基本原理与适用条件
    - 2.2.2 scipy的实现方式
    - 2.2.3 statsmodels的实现
  - ▼ 2.3 两样本t检验
    - 2.3.1 基本原理与适用条件
    - 2.3.2 scipy的实现方式
    - 2.3.3 statsmodels的实现
  - ▼ 2.4 配对样本t检验
    - 2.4.1 基本原理与适用条件
    - 2.4.2 scipy的实现方式
    - 2.4.3 statsmodels的实现

## 2.5 实战练习

## 2.5 实战练习

请考察北京、上海两地在2007年4月时的信心值是否有偏离基准值100。

请分北京、上海、广州三个城市来比较已婚人群和未婚人群的总指数、现状指数和预期指数是否有差异。

请自行完成CCSS\_pair数据中针对现状指数和预期指数变化情况的检验。

## 3 检验方法适用条件的考察

### 3.1 独立性的考察与应对策略

### 3.2 正态性的考察与应对策略

#### 3.2.1 用图形方法考察正态性

```
In [ ]: ccss.index1.plot.hist(bins = 18)
```

```
In [ ]: ccss.index1.plot.box()
```

#### 3.2.2 用假设检验考察正态性

```
scipy.stats.kstest(a vector like data, 'norm')
```