



首都师范大学

人工智能数学基础Python实践

第十二章 假设检验

张苗苗

信息工程学院

□ 在Python中实现以下操作：

- Z检验
- T检验
- 卡方检验

□ 所需的python库：

- `scipy`

名称	函数
正态性检验	<code>scipy.stats.shapiro</code> <code>scipy.stats.normaltest</code>
单独样本t检验	<code>scipy.stats.ttest_1sample</code>
两独立样本t检验	<code>scipy.stats.ttest_ind</code> <code>scipy.stats.ttest_ind_from_stats</code>
方差齐性检验	<code>scipy.stats.levene</code>
卡方检验	<code>scipy.stats.chisquare</code>

练习1-- 体检数据中的假设检验问题



数据集中包含多个人的体温、性别和心率数据，
试检验：

- (1) 体温的分布是否服从正态分布 **(正态性检验)**
- (2) 人类的平均体温是98.6 **(单总体T检验)**
- (3) 男性和女性的体温有差异吗？ **(两独立样本T检验 <-- 方差齐性检验)**

练习1-- 体检数据中的假设检验问题



(1) 读数据，观察数据，提取需要用的数据

```
data = pandas.read_csv(...)
observed_temperature = data[ 'Temperature' ]
female_temp = data.Temperature[data.Gender==2]
male_temp = data.Temperature[data.Gender==1]
```

(2) 正态性检验

```
scipy.stats.shapiro(observed_temperature) 或
scipy.stats.normaltest(observed_temperature)
```

(3) 单总体T检验

```
scipy.stats.ttest_1samp(observed_temperature, mu)
```

(4) 两独立样本T检验

```
scipy.stats.ttest_ind(female_temp, male_temp)
```

练习2-- 种族对求职是否有影响



数据集中包含多个求职者的数据，每个求职者的信息有55个变量描述，其中，变量race表示种族，b表示黑人，w表示白人；变量call表示面试是否录取，1表示录取，0表示未录取。试检验：

种族对求职是否有影响（卡方检验）

练习2-- 种族对求职是否有影响



(1) 读数据，提取实际观测值

```
data = pandas.io.stata.read_stata(...)
```

```
black = data[data.race == ' b' ]
```

```
white = data[data.race == ' w' ]
```

```
black_called =
```

```
black_not_called =
```

(2) 计算理论值

```
num_called = black_called + white_called
```

```
num_not_called = black_not_called + white_not_called
```

```
expected_called =
```

```
expected_not_called =
```

(3) 卡方检验

```
scipy.stats.chisquare(observed, expected)
```

在10块地上同时种植甲、乙两种作物，其产量服从正态分布，并且方差相同。结果计算得 $\bar{X} = 30.97$, $\bar{Y} = 21.79$, $S_x = 26.7$, $S_y = 12.1$ 。

试问这两种作物的产量有无明显差异？

从某中学随机抽取两个班，调查他们对待文理分科的态度，结果，甲班37人赞成，27人反对；乙班39人赞成，21人反对，试问这两个班对待文理分科的态度是否有显著差异？



首都师范大学

谢谢!