

## 2.4.2 scipy的实现方式

scipy.stats.ttest\_rel(

```
a, b : array_like  
nan_policy : {'propagate', 'raise', 'omit'}
```

)

In [ ]:

```
ccss_p.loc[:, ['index1', 'index1n']].describe()
```

In [ ]:

```
# 用相关分析确认配对信息是否的确存在  
ss.pearsonr(ccss_p.index1, ccss_p.index1n)
```

In [ ]:

```
ss.ttest_rel(ccss_p.index1, ccss_p.index1n)
```

In [ ]:

```
# 直接求出差值并进行单样本t检验  
ss.ttest_1samp(ccss_p.index1 - ccss_p.index1n, 0)
```

## 2.4.3 statsmodels的实现方式

statsmodels没有提供直接实现配对t检验的方法，但是可以有两个变通的实现方式

statsmodels.stats.ttest.paired : 提供两个界值点的单侧配对t检验结果  
求出差值，然后使用DescrStatsW.ttest\_mean()得到所需检验结果

In [ ]:

```
des = ws.DescrStatsW(ccss_p.index1 - ccss_p.index1n)  
des.ttest_mean()
```

## 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')
```

K-S检验，特点是比较严格，理论上可以检验任何分布。

```
scipy.stats.shapiro(a_vector_like_data)
```

Shapiro检验，专门用来检验正态分布。

```
scipy.stats.normaltest(a_vector_like_data)
```

D'Agostino and Pearson's 检验，基于峰度和偏度系数进行正态性检验。

```
scipy.stats.anderson(a_vector_like_data, dist='norm')
```

Anderson-Darling检验，为KS检验的改进。

In [ ]:

```
ss.kstest(ccss.index1, "norm")
```

In [ ]:

```
# 分组调用带参数的函数，用lambda实现
ks = lambda x: ss.kstest(x, "norm")
ccss.groupby('time').index1.apply(ks)
```

In [ ]:

```
# 直接进行无参数的分组检验
ccss.groupby('time').index1.apply(ss.shapiro)
```

## 3.3 方差齐性考察与应对策略

### 3.3.1 用图形方法考察方差齐性

In [ ]:

```
ccss.groupby('time').index1.plot.hist() # 默认设定图形会被叠加在一起无法阅读
```

In [ ]:

```
ccss.index1.hist(by = ccss.time) # 注意此处使用的是hist, 不是plot.hist
```

In [ ]:

```
ccss.boxplot(column = 'index1', by = 'time') # 注意参数的引用方法
```

In [ ]:

```
ccss.groupby('time').boxplot(column = 'index1') # 绘制箱图图组
```

### 3.3.2 用假设检验考察方差齐性

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

Bartlett's方差齐性检验，对数据有正态性要求。

`scipy.stats.levene(a, b, center = {'mean', 'median', 'trimmed'})`

Levene检验，在数据非正态时精度比Bartlett检验好，可调中间值的度量，默认median。

`scipy.stats.fligner(a, b, center='mean')`

Fligner-Killeen检验，非参检验，不依赖于分布。

In [ ]:

```
ss.bartlett(ccss.index1[ccss.time == 200704],
            ccss.index1[ccss.time == 200712],
            ccss.index1[ccss.time == 200812],
            ccss.index1[ccss.time == 200912]
            )
```

In [ ]:

```
ss.levene(ccss.index1[ccss.time == 200704],
           ccss.index1[ccss.time == 200712],
           ccss.index1[ccss.time == 200812],
           ccss.index1[ccss.time == 200912]
           )
```

## 3.4 实战练习

提取北京2007年4月的总指数、现状指数和预期指数，用假设检验方法和图形化方法考察其正态性，思考这两种方式的优缺点和使用场景，并比较不同检验方法的特点。

分别用图形方法和假设检验方法考察北京的总指数、现状指数和预期指数在四个时间点的方差齐性。

## 4 单因素方差分析

### 4.1 基本原理与适用条件