

Python实现描述性统计的内容

集中趋势

离散程度

分布形状

Python实现描述性统计的内容

集中趋势

```
import numpy as np
import pandas as pd
df = pd.read_excel('data/week3_monthly_income.xlsx')
data = df['月收入']
```

- 众数

```
# 众数
# 方法一
print('众数为: ', data.mode().iloc[0])

# 方法二
from collections import Counter
counts = Counter(data)
print('众数为: ', counts.most_common(1)[0][0])

# 方法三
from scipy.stats import mode
mode_num = mode(data)
print('众数为: ', mode_num[0][0])
```

- 中位数

```
# 中位数
# 方法一
print('中位数为: ', data.median())
# 方法二
print('中位数为: ', np.percentile(data, 50))
# 方法三
print('中位数为: ', data.quantile(.50))
```

- 分位数

```
# 分位数
# 方法一
print('下四分位数为: ', np.percentile(data, 25))
print('上四分位数为: ', np.percentile(data, 75))
# 方法二
print('下四分位数为: ', data.quantile(.25))
print('上四分位数为: ', data.quantile(.75))
```

- 平均数
 - 算术平均数

```
# 平均数
# 算术平均数
print('算术平均数为: ', data.mean())
```

- 几何平均数

```
# 几何平均数
import math
s = 1
for i in data:
    s *= i
print('几何平均数为: ', math.pow(s, 1/len(data)))
```

离散程度

- 异众比率

```
# 异众比率
# 众数
from collections import Counter
count = Counter(data)
modeCount = count.most_common(1)[0][1]
totalCount = len(data)
# 异众比率
ratio = (totalCount - modeCount) / totalCount
print(ratio)
```

- 四分位差

```
# 四分位差
# 方法一
print('四分位差为: ', np.percentile(data, 75) - np.percentile(data, 25))
# 方法二
print('四分位差为: ', data.quantile(.75) - data.quantile(.25))
```

- 极差

```
# 极差
print('极差为: ', data.max() - data.min())
```

- 平均差

```
# 平均差
meanData = data.mean()
s = 0
for i in data:
    s += (abs(i - meanData))
print('平均差为: ', s/len(data))
```

- 方差

```
# 方差
print('方差为: ', np.var(data))
```

- 标准差

```
# 标准差
print('标准差为: ', np.std(data))
```

分布形状

- 偏度

```
# 偏度
from scipy import stats
print('偏度为: ', stats.skew(data))
```

- 峰度

```
# 偏度
from scipy import stats
print('偏度为: ', stats.skew(data))
```