

1. 数据描述

```
import pandas as pd
import scipy.stats as ss
```

```
df = pd.read_excel('data.xlsx', index_col=0)
data = df.drop(df.columns[0], axis=1)
columns = data.columns

data.head()
```

	Salary income	Household income	Property income	Transfer income
No				
1	4524.25	1778.33	588.04	455.64
2	2720.85	2626.46	152.88	79.64
3	1293.50	1988.58	93.74	105.81
4	1177.94	1563.52	62.70	86.49
5	504.46	2223.26	73.05	188.10

统计数据基本情况

- 总数
- 均值
- 标准差
- 最小值
- 4分位数
- 最大值

```
data.describe()
```

	Salary income	Household income	Property income	Transfer income
count	31.000000	31.000000	31.000000	31.000000
mean	1394.466452	1822.891613	115.492581	178.723548
std	1313.328388	484.057977	128.721523	154.307114
min	195.510000	774.600000	16.810000	52.660000
25%	574.230000	1523.245000	38.720000	98.260000
50%	954.890000	1778.330000	62.700000	129.830000
75%	1365.535000	2193.375000	149.395000	198.760000
max	6159.700000	2789.400000	588.040000	855.950000

计算变异系数（偏差标准偏差与平均值的比值）

```
ss.variation(data)
```

```
array([ 0.92649919,  0.26122595,  1.09641978,  0.84934464])
```

计算偏度

```
data.skew()
```

```
Salary income      2.242446
Household income   -0.047204
Property income     2.419086
Transfer income     3.206948
dtype: float64
```

计算峰度

```
df.kurt()
```

```
Salary income      5.439312
Household income   -0.546945
Property income     6.348023
Transfer income    12.388402
dtype: float64
```

2. 关于SPSS箱线图胡须位置

从箱图延伸出的 T 形条称为内围或细线。这些条延伸至箱图高度的 1.5 倍，或者如果个案/行有一个在该范围内的值，则延伸至最小或最大值。 [SPSS 官方教程1](#)

所以spss的箱线图的胡须是从箱线图的顶部或底部向外延伸1.5倍的箱体长度。