

## 任务 2: Pandas 进阶、统计学算法基础的 Pandas 实现

- python 中的 lambda, Pandas 在进行排序或筛选时会经常用到
- Pandas 中的 Where 查询
- Pandas 中的透视表
- Pandas 中的字符串处理
- Pandas 中的时序表
- apply()方法
- 一维线性回归

参考: <https://liaocy.net/2020/05/02-statistics-formula/>

参考:

[http://www.360doc.com/content/20/0225/07/6348482\\_894631222.shtml](http://www.360doc.com/content/20/0225/07/6348482_894631222.shtml)

Pandas 中 Series 或 DataFrame 自带的所有计算函数, 具体可以查询:

<https://pandas.pydata.org/pandas-docs/stable/reference/series.html#computations-descriptive-stats>

用 Pandas 实现以下统计算法, 并且要求必须从公式层面掌握:

- 均值
- 中位数
- 众数和异众比率
- 极差
- 方差和标准差
- 协方差
- 一维离散数据概率分布直方图
- 正态分布的拟合

完成任务：

## 一、根据 **Wind.data** 完成以下任务

1. 导入 Wind 数据，并展示表头
2. 设法解决违规日期（61 年）的问题，将其改为 1961 年
3. 将日期设为 datetime64 索引
4. 统计缺失值个数
5. 计算每个地区的风速平均值
6. 创建一个名为 loc\_stats 的数据框去计算并存储每个 location 的风速最小值，最大值，平均值和标准差
7. 创建一个名为 day\_stats 的数据框去计算并存储每天的风速最小值，最大值，平均值和标准差
8. 对于每一个 location，计算一月份的平均风速
9. 对于数据，分别以年、月为频率取样

## 二、

1. 使用 `second_cars_info.csv` 制作一个查询引擎，可以通过输入指定的参数获取对应条目的结果。
2. 使用 `US_Crime_Rates_1960_2014.csv`，根据 1991-2014 年的数据制作一条回归曲线，预测 2019 年的各个犯罪条目的犯罪率。