

# 分析数值数据

---

## 数值分析的四个方面

分析数值数据有四个主要方面:

1. Center 集中趋势测量
2. Spread 离散程度测量
3. Shape s数据的形状
4. Outliers 异常值

## 分析分类数据

分类数据的分析方法通常是查看落入每个组的独立个体的数量或比例。

如果我们在看狗的品种，我们会关心每个品种有多少只狗，或者每个品种的狗的比例如何。

## 集中趋势测量

---

集中趋势测量的方式

1. Mean 均值
2. Median 中位数
3. Mode 众数

## 均值

均值 在数学中通常称为平均数或预期值。我们通过将所有值相加，然后除以数据集中所有测量值的个数来计算均值。

## 中位数

中位数 将我们的数据分为两部分，一半低于它，一半高于它。我们在此视频中发现，如何计算中位数取决于我们有偶数个还是奇数个观察值。

### 奇数个值的中位数

如果我们有奇数个观察值，中位数直接是中间的那个数字。

如我们有 7个观察值并按从小到大排列，则中位数是第四个数。如果我们有9个观察值，则中位数是第五个数。

### 偶数个值的中位数

如果我们有偶数个观察值，中位数是中间两个值的平均值。例如，如果我们有8个观察值并按从小到大排列，则计算第四和第五个值的平均值。

**计算中位数,我们必须首先对值排序**

## 众数

众数 指一组数据中出现次数最多的数据值。  
一个数据集中可能有多个众数，也可能没有众数。

### 无众数

如果数据集中的所有值出现的频数相同，则不存在众数。如果我们有一组数据集：

1, 1, 2, 2, 3, 3, 4, 4

则没有众数，因为所有观察值发生的次数相同。

### 多个众数

如果两个（或多个）数字出现的次数都是最多的，则有多个众数。如果我们有一组数据集：

1, 2, 3, 3, 3, 4, 5, 6, 6, 6, 7, 8, 9

其中有两个众数 3和6，因为这两个值都出现了三次，出现频率最高，而其他的值都只出现了一次。