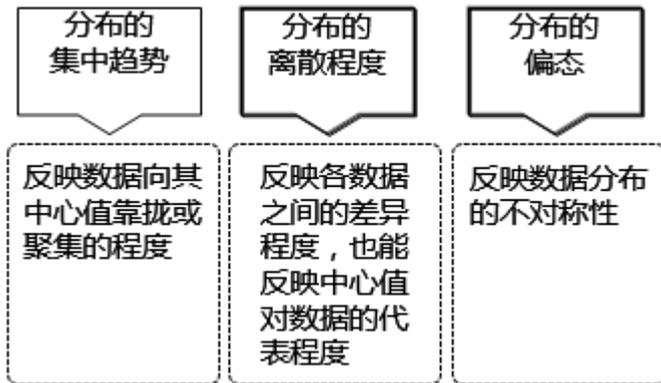


## 第二十四章 描述统计

在描述统计中可通过统计量描述数据的分布特征  
对于数据分布特征的测度分为三个方面



- 第一节 集中趋势的测度
- 第二节 离散程度的测度
- 第三节 分布形态的测度
- 第四节 变量间的相关分析

### 第一节 集中趋势的测度

均值★★★★★

中位数★★★★★

众数★★★★★

均值、中位数和众数的比较及适用范围★★★

#### 集中趋势的测度

集中趋势是指一组数据向某一中心值靠拢的程度，它反映了一组数据中心点的位置所在。  
集中趋势的测度也就是寻找数据一般水平的代表值或中心值。

均值含义	均值也叫平均数，就是数据组中所有数值的总和除以该组数值的个数。
计算公式	$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$

#### 示例

【教材例题】某售货小组有 5 名营业员，元旦一天的销售额分别为 520 元、600 元、480 元、750 元和 500 元，求该日平均销售额。

【计算过程】  
平均销售额 =  $\frac{520 + 600 + 480 + 750 + 500}{5} = 570$  元

地位	均值是集中趋势最主要的测度值，它是一组数据的重心所在，解释了一组数据的平均水平。
适用范围	主要适用于数值型数据 但不适用于分类数据和顺序数据

特点	易受极端值的影响，极端值的出现，会使平均数的真实性受到干扰。
----	--------------------------------

【单选题】集中趋势最主要的测度值是（ ）。

- A.众数
- B.中位数
- C.均值
- D.方差

【答案】C

【解析】均值也叫平均数，是集中趋势中最主要的测度指标。

【单选题】某城市 2014 年 4 月空气质量检测结果中，随机抽取 6 天的质量指数进行分析。样本数据分别是：30、40、50、60、80 和 100，这组数据的平均数是（ ）。

- A.50
- B.55
- C.60
- D.70

【答案】C

【解析】 $(30+40+50+60+80+100) \div 6=60$

## 考点 2 中位数

中位数 含义	把一组数据按从小到大或从大到小的顺序进行排列，位置居中的数值叫做中位数，用 Me 表示。 中位数将数据分为两部分，其中一半的数据小于中位数，另一半数据大于中位数。
-----------	--

### 计算思路

数据排序→确定中位数的位置→确定中位数的具体数值

n 为数据的个数

(1) 当 n 为奇数：中位数位置是  $\frac{n+1}{2}$ ，该位置所对应的数值就是中位数数值

(2) 当 n 为偶数：中位数位置是介于  $\frac{n}{2}$  和  $(\frac{n}{2}+1)$  之间，中位数数值就是  $\frac{n}{2}$  所对数值和  $(\frac{n}{2}+1)$  所对应数值的平均数。

### 示例 1

【教材例题】某地级市下辖 9 个县，每个县的面积如下，计算中位数：

1455 2019 912 1016 1352 1031 2128 1075 2000

【计算思路】

(1) 先排序：

912 1016 1031 1075 1352 1455 2000 2019 2128

(2) 确定中位数的位置  $(9+1) \div 2=5$ ，

(3) 该位置所对应的数值就是中位数数值中位数 1352

### 示例 2

【教材例题】某地级市下辖 10 个县，每个县的面积如下，计算中位数：

912 1000 1016 1031 1075 1352 1455 2000 2019 2128

【计算思路】

(1) 先排序：题目中数据已经从小到大排好序

(2) 确定中位数的位置  $(10+1) \div 2=5.5$

(3) 即第 5 位数和第 6 位数所对应数值的平均数就是中位数, 即:

$$Me = \frac{1075 + 1352}{2} = 1213.5$$

### 适用范围

主要用于顺序数据和数值型数据

但不适用于分类数据

### 特点

中位数是一个位置代表值, 不受极端值的影响, 抗干扰性强, 尤其适用于收入类偏斜分布的数值型数据。

**【单选题】** 2010 年某省 8 个地市的财政支出 (单位: 万元) 分别为: 59000 50002 65602 66450 78000 78000 78000 132100 这组数据的中位数是 ( ) 万元。

- A.78000
- B.72225
- C.66450
- D.75894.25

**【答案】** B

**【解析】** 把一组数据按从小到大的顺序进行排列, 位置居中的数值叫做中位数。把题干中的数据按顺序排列就是 50002, 59000, 65602, 66450, 78000, 78000, 78000, 132100。

由于所给数据是 8 个, 所以中位数的位置是第 4 个和第 5 个数据的平均数。

中位数就是  $(66450+78000) \div 2=72225$ 。

**【单选题】** 在对数据集中趋势的测度中, 适用于偏斜分布的数值型数据是 ( )。

- A.中位数
- B.均值
- C.标准差
- D.方差

**【答案】** A

**【单选题】** 下列数据特征测度中, 是用于反映偏斜分布数值型数据集中趋势的是 ( )。

- A.离散系数
- B.方差
- C.标准差
- D.中位数

**【答案】** D

### 考点 3 众数

众数

#### 含义

众数是一组数据中出现次数 (频数) 最多的变量值。

#### 示例

某能源公司有 9 个分公司, 每个分公司的主营产品分别是:

煤制品、有机化工原料、火电、煤制品、热力、电解铝、火电、

煤制品、煤制品, 则该能源公司分公司主营产品众数为煤制品。

众数

### 适用范围

适于描述分类数据和顺序数据的集中趋势，而有些情况下可能出现双众数、多众数或者没有众数，难以描述数据的集中位置，不适用于定量数据。

**【单选题】** 2010年某省8个地市的财政支出（单位：万元）分别为：59000 50002 65602 66450 78000 78000 78000 132100 这组数据的中位数和众数分别是（ ）万元。

- A.78000 78000
- B.72225 78000
- C.66450 132100
- D.75894 25132100

**【答案】** B

**【解析】** 本题先选择众数，可以排除CD。再确定中位数，由于所给数据是8个，所以中位数的位置是第4个和第5个数据的平均数。 $(66450+78000)/2=72225$

均值、中位数和众数的比较及适用范围					
	指标 (测得值)	是否受极端值影响	数值型数据 (定量数据)	分类数据	顺序数据
测度数据集中趋势	均值	受影响	适用	不适用	不适用
	中位数	不受影响	适用	不适用	适用
	众数	不受影响	不适用	适用	适用

均值、中位数和众数的比较及适用范围		
指标	优点	缺点
均值	能够充分利用数据的全部信息，均值大小受每个观测值的影响，比较稳定	易受极端值的影响，如果观测值中有明显的极端值，则均值的代表性较差
中位数	不受极端值的影响	没有充分利用数据的全部信息，稳定性差于均值，优于众数
众数	不受极端值的影响，尤其是分布明显呈偏态时，众数的代表性更好	没有充分利用数据的全部信息，缺乏稳定性，而且可能不唯一

**(2019)** 关于数据集中趋势测度的说法，正确的有（ ）。

- A.集中趋势反映数据分布的不对称性
- B.集中趋势反映了一组数据中心点的位置所在
- C.中位数适用于分类数据集中趋势的测度
- D.集中趋势的测度是寻找数据水平的代表值或中心值
- E.集中趋势是指一组数据向某一中心值靠拢的程度

**【答案】** BDE

**【解析】** 集中趋势是指一组数据向某一中心值靠拢的程度，它反映了一组数据中心点的位置所在。集中趋势的测度也就是寻找数据水平的代表值或中心值。中位数不适用于分类变量，适于顺序变量和定量变量，特别是分布不对称的数据。

**(2020)** 关于众数的说法，正确的有（ ）。

- A.在定量数据中，可能出现多众数和无众数的情况
- B.众数是算数平均数的一种
- C.众数是指一组数据中出现次数最多的变量值
- D.众数适用于描述分类数据的集中趋势
- E.众数适用于描述顺序数据的集中趋势

**【答案】** ACDE

**【解析】**在定量数据中，可能出现多众数和无众数的情况，所以众数不适用于定量变量。众数是位置平均数，B选项错误。众数适用于分类数据和顺序数据。

**（2022）**下列统计数据适用于数值统计数据的有（ ）。

- A.标准差
- B.方差
- C.众数
- D.中位数
- E.均值

**【答案】** ABDE

**【解析】**均值、方差和标准差仅适用数值型数据。中位数可以适用于顺序数据，也可以适用于数值型数据。众数适用于分类数据和顺序数据，不适用于数值型数据。