

第二十四章
描述统计
框架导航



总论

数据分布特征的测度

- 在描述统计中，可以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主要分 为三个方面：
1. 分布的**集中趋势**，反映各数据向其中心值**靠拢**或**聚集**的**程度**；
 2. 分布的**离散程度**，反映各数据之间的**差异程度**，也能反映中心值对数据的代表程度；
 3. 分布的**偏态**，反映数据**分布的不对称性**。

测 度	方 法
集 中 趋 势	均 值 中 位 数 众 数
离 散 程 度	方 差 标 准 差 离 散 系 数
偏 态	偏 态 系 数 标 准 分 数

【2025年·多选】能表示数据集中趋势的是（ ）。

- A. 方差
- B. 均值
- C. 标准差
- D. 中位数
- E. 众数

答案： BDE

解析： 集中趋势的测度指标包括均值、中位数、众数，故BDE选项正确。A选项方差、C选项标准差是反映数据离散程度的指标。

第一节

集中趋势的测度

- 1. 均值
- 2. 中位数
- 3. 众数
- 1. 均值

均值也叫作平均数，就是数据组中所有数值的总和除以该组数值的个数。设一 组数据为

$X_1, X_2, \dots, X_n:$

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n}$$

某售货小组有5名营业员，元旦全天的销售额分别为520元、600元、480元、750元 和500元，求该日平均销售

额。

$$\bar{X} = \frac{520 + 600 + 480 + 750 + 500}{5} = 570 \text{ (元)}$$

项目	均值
内涵	均值是 集中趋势最主要的测度值 ，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	适用于 定量变量 。
优点	能够充分利用数据的全部信息，比较稳定。
缺点	易受极端值的影响 。

2. 中位数

把一组数据按从小到大或从大到小的顺序进行排列，位置居中的数值叫作中位数，用M表示。

$$M_e = \begin{cases} X_{(\frac{n+1}{2})} & n \text{为奇数} \\ \frac{1}{2} [X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}] & n \text{为偶数} \end{cases}$$

地级市下辖9个县，每个县的面积如下（单位：平方千米），计算该市下辖县面积的中位数：

1455, 2019, 912, 1016, 1352, 1031, 2128, 1075, 2000

1.排序

912, 1016, 1031, 1075, 1352, 1455, 2000, 2019, 2128

2.中间位置确定

$$M_e = X_{(\frac{9+1}{2})} = 1352 \text{ (平方千米)}$$

项目	中位数
内涵	把一组数据按 从小到大或从大到小的顺序进行排列 ，位置居中的数值叫作中位数，用表示。

适用范围	适用于 顺序变量和定量变量 ，特别是 分布不对称 的数据。
优点	不受极端值的影响。
缺点	没有充分利用数据的全部信息， 稳定性差于均值，优于众数 。

3. 众数

众数是指一组数据中出现次数（频数）最多的变量值。

【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。

众数：煤制品

项目	众数
内涵	一组数据中出现次数（频数）最多的变量值。
适用范围	适用于 分类和顺序变量 。
优点	不受极端值的影响。
缺点	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。

项目	均值	中位数	众数
适用范围	适用于 定量变量 。	适用于 顺序变量和定量变量 ，特别是 分布不对称 的数据。	适用于 分类和顺序变量 。
优点	能够充分利用数据的全部信息，比较稳定。	不受极端值的影响。	不受极端值的影响。
缺点	易受极端值的影响 。	没有充分利用数据的全部信息， 稳定性差于均值，优于众数 。	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。

【2021年·单选】2020年某餐饮企业12个加盟店的年营业收入（单位：万元）分别为：1211，1255，1257，1323，1345，1422，1478，1488，1573，1623，1633，1822。这组数据的中位数是（ ）万元。

- A. 1422
- B. 1500
- C. 1578
- D. 1450

答案：D

解析：把一组数据按从小到大或从大到小的顺序进行排列，位置居中的数值叫作中位数。根据题干可知共有12个数据，中位数为第6和第7个位置数据的平均数，即中位数=（1422+1478）/2=1450。

【2025年·单选】某超市某类商品价格数据的众数为25元，则可以判断出（ ）。

- A. 该组价格数据为对称分布
- B. 该组价格数据中位数为25元
- C. 该类商品平均价格为25元
- D. 该组数据中25元这一价格出现的次数最多

答案： D

解析：众数是一组数据中出现次数最多的变量值，题干明确某类商品价格数据的众数为25元，直接可得出25元这一价格出现的次数最多，故D选项正确。

【2025年·单选】下列数据特征中测度不受极端值影响的是（ ）。

- A. 方差
- B. 标准差
- C. 中位数
- D. 均值

答案： C

解析：中位数是位置平均数，不受极端值的影响，故C选项正确。A、B、D均受极端值影响。

【2024年·单选】下列选项中，适合用众数测度集中趋势的是（ ）。

- A. 销售额
- B. 工资性收入
- C. 企业利润
- D. 城乡属性

答案： D

解析：众数适用于分类数据，城乡属性属于分类变量，故D正确。A、B、C为数值型数据。

【2025年·多选】关于中位数的说法，正确的是（ ）。

- A. 中位数适合分布不对称的定量数据
- B. 中位数不受极端值的影响
- C. 中位数是一组数据中出现次数最多的数
- D. 中位数利用了数值全部信息
- E. 中位数适用于分类数据

答案： AB

解析：中位数适合偏态分布的定量数据，不受极端值影响，故AB选项正确。C选项是众数的定义；D选项错误，

中位数未利用全部数值信息；E选项错误，中位数不适用于分类数据。

第二节

离散程度的测度

1. 方差

2. 标准差

3. 离散系数

1. 方差

方差是数据组中各数值与其均值离差平方的平均数，它能较好地反映出数据的离散程度。方差**越小**，说明数据值与均值的平均距离越小，**均值的代表性越好**。

(1) 总体方差：

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N} \quad \text{或} \quad S^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N - 1}$$

(2) 样本方差：

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n - 1}$$

某售货小组有5名营业员，元旦全天的销售额分别为520元、600元、480元、750元 和500元，求该日平均销售额。

$$\bar{X} = \frac{520 + 600 + 480 + 750 + 500}{5} = 570 \text{ (元)}$$

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n - 1} \\ &= \frac{(520 - 570)^2 + (600 - 570)^2 + (480 - 570)^2 + (750 - 570)^2 + (500 - 570)^2}{5 - 1} \\ &= 12200 \end{aligned}$$

2. 标准差

方差是反映数据离散程度的重要测度指标，但是其单位是原数据单位的平方，没有解释意义。我们经常使用

标准差来测度数据的离散程度，标准差即方差的平方根。

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

某售货小组有5名营业员，元旦全天的销售额分别为520元、600元、480元、750元 和500元，求该日平均销售额。

$$\begin{aligned} s^2 &= \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n - 1} \\ &= \frac{(520 - 570)^2 + (600 - 570)^2 + (480 - 570)^2 + (750 - 570)^2 + (500 - 570)^2}{5 - 1} \\ &= 12200 \end{aligned}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} = \sqrt{12200} \approx 110.45 \text{ (元)}$$

3. 离散系数

离散系数也称变异系数或标准差系数，即标准差与均值的比值，主要用于不同类别数据离散程度的比较，记为CV。

$$CV = \frac{s}{\bar{X}}$$

某售货小组有5名营业员，元旦全天的销售额分别为520元、600元、480元、750元 和500元，求该日平均销售额。

$$\bar{X} = \frac{520 + 600 + 480 + 750 + 500}{5} = 570 \text{ (元)}$$

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} = \sqrt{12200} \approx 110.45 \text{ (元)}$$

$$CV = \frac{s}{\bar{X}} = \frac{110.45}{570} \approx 0.19$$

项目	方差	标准差	离散系数
内涵	方差是数据组中各数值与其均值离差平方的平均数。	标准差即方差的平方根。 $s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$	离散系数也称变异系数或标准差系数，即标准差与均值的比值，记为 CV。 $CV = \frac{s}{\bar{X}}$
	方差的单位是原数据单位的平方，没有解释意义。	标准差不仅能度量数值与均值的平均距离，还与原始数值具有相同的计量单位。	主要用于不同类别数据离散程度的比较。
注意	标准差与方差计算简单、有较好的数学性质，是应用最广泛的统计离散程度的测度方差。		

【2025年·单选】标准差与方差的关系是（ ）。

- A. 标准差是方差与样本量的乘积
- B. 方差是标准差与样本量的乘积
- C. 方差是标准差的平方
- D. 标准差是方差的平方

答案：C

解析：方差是标准差的平方，故C正确。

【2025年·多选】关于离散系数的说法，正确的有（ ）。

- A. 离散系数适用于分类数据
- B. 离散系数不受观测水平的影响
- C. 离散系数不受观测单位的影响
- D. 离散系数适用于顺序数据
- E. 离散系数不受极端值的影响

答案：BC

解析：离散系数（变异系数）不受观测水平影响（B）和观测单位影响（C）。A错误，仅适用于数值型数据；D错误，不适用于顺序数据；E错误，受极端值影响。

【2023年·单选】一组数据的方差为100、均值为40，则这组数据的离散系数为（ ）。

- A. 4
- B. 0.25
- C. 2.5
- D. 0.4

答案：B

解析：离散系数也称为变异系数或标准差系数，即标准差与均值的比值： $CV=10/40=0.25$ 。

总结

	指标 (测度值)	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
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
离散趋势	方差	受影响	适用	不适用	不适用
	标准差				
	离散系数				