



第二十四章

描述统计



第一节 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水平的代表值或中心值。



第一节 集中趋势的测度

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 $X_1, X_2, \dots, X_n$ ，平均数 $\bar{X}$ 的计算公式为：

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



第一节 集中趋势的测度

【例】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。

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

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。它主要适用于数值型数据，但不适用于分类和顺序数据。此外，均值容易受到极端值的影响，极端值会使得均值向极大值或极小值方向倾斜，使得均值对数据组的代表性减弱。



第一节 集中趋势的测度

2、中位数

(1) 定义

把一组数据按从小到大或从大到小的顺序进行排列，位置居中的数值叫作**中位数**，用 M_e 表示。中位数将数据分成两部分，其中一半数据小于中位数，另一半数据大于中位数。



第一节 集中趋势的测度

(2) 公式

设一组数据为 $X_1, X_2, \dots, X_n$, 按从小到大顺序为 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, 则中位数为:

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



第一节 集中趋势的测度

【例】某地级市下辖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：

1 455 2 019 912 1 016 1 352 1 031 2 128 1 075

2 000

第一步，将上面的数据按从小到大的顺序进行排列：

912 1 016 1 031 1 075 1 352 1 455 2 000 2 019

2 128



第一节 集中趋势的测度

第二步，判断奇数还是偶数，计算中位数位置：

中位数位置 = $(9+1) \div 2 = 5$ ，中位数为1 352，

即 $M_e = 1\ 352$ （平方千米）

在上例中，由于行政区划调整，临市的一个面积为1 000平方千米的县划归该市。行政区划调整后，该市现在下辖10个县，该市下辖县的面积（单位：平方千米）从小到大依次为：

912 1 000 1 016 1 031 1 075 1 352 1 455 2 000 2 019

2 128

计算行政区划调整后该市下辖县面积的中位数：

中位数位置 = $\frac{1075+1352}{2} = 1213.5$ 中位数为
 $M_e = 1\ 213.5$ （平方千米）



第一节 集中趋势的测度

(3) 中位数是一个位置代表值，主要用于顺序数据和数值型数据，但不适用于分类数据。中位数的优点是不受极端值的影响，抗干扰性强，尤其适于收入这类偏斜分布的数值型数据。



第一节 集中趋势的测度

3、众数

众数是指一组数据中出现次数（频数）最多的变量值。

例如，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。

众数适用于描述**分类数据**和**顺序数据**的集中趋势。而在定量数据中，可能出现多众数和无众数的情况，因此众数**不适用于描述定量数据的集中位置**。



第一节 集中趋势的测度

4、均值、中位数和众数的比较及适用范围

集中趋势指标	适用数据类型	是否受 极端值影响	利用数据的 全部信息
均值	定量变量	受	充分利用
中位数	顺序变量、定量变量	不受	没有
众数	分类变量、顺序变量	不受	没有



第一节 集中趋势的测度

【例-单选题】下列指标中，用于描述数据集中趋势，并且易受极端值影响的是（ ）。

- A. 算术平均数
- B. 中位数
- C. 众数
- D. 方差

答案：A

解析：BC项中位数和众数虽然可以描述数据集中趋势，但都不受极端值的影响，D项方差是描述数据离散程度，只有A项算术平均数用于描述数据集中趋势，并且易受极端值影响。



第一节 集中趋势的测度

【例-单选题】2020年某省8个地市的财政支出（单位：万元）分别为：

59 000 50 002 65 602 66 450 88 000 88 000 88 000 132 100，这组数据的众数是（ ）万元。

A. 88 000

B. 72 225

C. 66 450

D. 75 894

答案：A

解析：众数是指一组数据中出现次数（频数）最多的变量值。