初级经济师
经济基础知识
教材精讲班

第二十一章 数据特征的测度

第二十一章 数据特征的测度		
考点 1	集中趋势的测度	★★★
考点 2	离散程度的测度	★★

考点 1 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倾向。

一、众数、中位数		
	众数	中位数
计算方法	一组数据中出现频数最多的那个数值，用 M_0 表示	把一组数据按从小到大的顺序进行排列，位置居中的数值叫作中位数，用 M_e 表示。
		$M_e=\begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{当n为奇数时} \\ \frac{1}{2}\left(X_{\frac{n}{2}}+X_{\frac{n}{2}+1}\right) & \text{当n为偶数时} \end{cases}$
特点	是一个位置代表值，不受极端值的影响，抗干扰性强	
适用范围	适用于品质数据、数值型数据	主要用于顺序数据，也适用于数值型数据，但不适用于分类数据。

【例如】一家连锁超市的 10 个分店某月的销售额（单位：万元）分别为：61，65，73，78，80，80，80，80，96，97。

这 10 个分店月销售额的众数为 M0=80（万元）

【例如】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：

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

首先，将上面的数据排序：912，1016，1031，1075，1352，1455，2000，2019，2128

中位数位置=（9+1）÷2=5，中位数为 1352，即 Me=1352（平方千米）

【单选题】下列属于特征测度中，属于位置平均数的是（　　）。

A.几何平均数

B.算数平均数

C.中位数

D.极差

【答案】C

三、算术平均数	
定义	是全部数据的算术平均，又称均值，用　表示。
特点	（1）是集中趋势最主要的测度值。

	（2）易受极端值的影响。 （3）受到两个因素的影响：一个是各组数值的大小，另一个是各组分布频数的多少。
适用范围	主要适用于数值型数据，但不适用于品质数据。

三、算术平均数 两种算数平均数的比较		
	简单算术平均数	加权算术平均数
计算方法	设一组数据为 X_1，X_2，…，X_n，公式为：	设原始数据被分成 k 组：各组组中值为 X_1，X_2，…，X_k，各组的频数分别为 f_1，f_2，…，f_k，公式为：
	$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}=\frac{\sum_{i=1}^{n}X_i}{n}$	$\overline{X}=\frac{X_1f_1+X_2f_2+\cdots+X_kf_k}{f_1+f_2+\cdots+f_k}=\frac{\sum_{i=1}^{k}X_if_i}{\sum_{i=1}^{k}f_i}$
	处理未分组的原始数据	处理经分组整理的数据

三、算术平均数

【例如】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，求该日每名营业员的平均销售额。

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

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

【例如】某市商业企业协会根据 100 个会员样本，整理出一年销售额分布资料：

销售额（万元）	组中值 Xi	商业企业数 f_i	X_if_i
100-150	125	4	500
150-200	175	16	2800
200-250	225	40	9000
250-300	275	28	7700
300-350	325	10	3250
350-400	375	2	750
合计	—	100	24000

计算年平均销售额。

首先确定组中值，计算结果在表中列出。代入公式得：

$$\overline{X}=\frac{\sum_{i=1}^{k}X_if_i}{\sum_{i=1}^{k}f_i}=\frac{24000}{100}=240\text{（万元）}$$

1. n 个观察值连乘积的 n 次方根就是几何平均数。

2. 简单几何平均数的计算。

设一组数据为X_1，X_2，…，X_n，且大于0，$\overline{X_G}$表示几何平均数，则：

主要用途： $\overline{X_G}=\sqrt[n]{X_1 \times X_2 \times \cdots \times X_n}=\sqrt[n]{\prod_{i=1}^{n} X_i}$

（1）对比率、指数等进行平均。
（2）计算平均发展速度。

【例如】某型号钻头的生产，需经过 6 道不同的加工工序，各道工序的合格率如下表所示，计算平均合格率。各道加工工序合格率用几何平均数的方法进行计算，得：

工序名称	合格率（%）
冲料	98.2
料废	97.5
车工	97.0
加热	96.6
扫槽	95.5
接柄	95.0

$$\overline{X_G}=\sqrt[n]{\prod_{i=1}^{n} X_i}$$

$$=\sqrt[6]{98.2\% \times 97.5\% \times \cdots \times 95.0\%}$$

$$=96.63\%$$

<table>
<tr><th colspan="6">【总结】</th></tr>
<tr><td rowspan="2"></td><td rowspan="2">指标</td><td rowspan="2">极端值</td><td colspan="2">品质数据</td><td rowspan="2">数值型数据</td></tr>
<tr><td>分类数据</td><td>顺序数据</td></tr>
<tr><td rowspan="2">位置
平均数</td><td>众数</td><td rowspan="2">不受影响</td><td>适用</td><td>适用</td><td>适用</td></tr>
<tr><td>中位数</td><td>不适用</td><td>适用</td><td>适用</td></tr>
<tr><td rowspan="2">数值
平均数</td><td>算术平均数</td><td rowspan="2">受影响</td><td>不适用</td><td>不适用</td><td>适用</td></tr>
<tr><td>几何平均数</td><td>不适用</td><td>不适用</td><td>适用</td></tr>
</table>

考点 2 离散程度的测度

一、离散程度的测度
1. 离散程度是指数据之间差异程度或频数分布的分散程度。
2. 离散程度和集中趋势是两个同样重要的数据分布特征。集中趋势的测度值是对数据一般水平的概括性变量，它对一组数据的代表程度，取决于该组数据的离散水平。
3. 数据的离散程度越大，集中趋势的测度值对该组数据的代表性就越差；离散程度越小，其代表性就越好。

<table>
<tr><th colspan="3">二、极差、标准差和方差</th></tr>
<tr><td></td><td>极差</td><td>标准差</td></tr>
<tr><td>含义</td><td>是最简单的变异指标，是总体或分布中最大的标志值与最小的标志值之差，又称全距，用 R 表示。</td><td>标准差是总体所有单位标志值与其平均数离差之平方的平均数的平方根，用 σ 表示。</td></tr>
</table>

计算	R＝Xmax－Xmin	$\sigma=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}}$（用于未整理的原始数据） $\sigma=\sqrt{\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2 f_i}{\sum_{i=1}^{n} f_i}}$（用于分组数据）

注意

1.极差反映的是变量分布的变异范围或离散程度；计算简便，含义直观，运用方便；但它不能反映其间的变量分布情况，还易受极端值的影响。

2.方差：方差是标准差的平方。

计算公式：

$$\sigma^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}$$（用于未整理的原始数据）

$$\sigma^2=\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2 f_i}{\sum_{i=1}^{n} f_i}$$（用于分组数据）

三、离散系数	
含义	离散系数通常是就标准差来计算的，因此也称标准差系数； 它是一组数据的标准差与其相应的算术平均数之比，是测度数据离散程度的相对指标。
目的	为了消除变量值水平高低和计量单位不同对离散程度测度值的影响。
计算公式	一组数据的标准差与其相应的算术平均数之比。
应用	主要是用于比较对不同组别数据的离散程度。 离散系数大的说明数据的离散程度也就大，离散系数小的说明数据的离散程度也就小。

【单选题】某学校学生的平均年龄为 20 岁，标准差为 3 岁；该校教师的平均年龄为 38 岁，标准差为 3 岁。比较该校学生年龄和教师年龄的离散程度，则（　）。

A.学生年龄和教师年龄的离散程度相同

B.教师年龄的离散程度大一些

C.教师年龄的离散程度是学生年龄离散程度的 1.9 倍

D.学生年龄的离散程度大一些

【答案】D

【解析】本题考查离散系数。平均值不同的情况下，用离散系数比较离散程度。学生年龄的离散系数＝3/20×100%＝15%。教师年龄的离散系数＝3/38×100%＝7.89%。离散系数大的说明数据的离散程度就大。