第四章 数据分布特征的描述

教学目的

通过本章的学习，了解数据分部的两种趋势及所用指标，理解各种指标的特点和应用场合并熟练掌握其计算方法，能作简单的分析。

本 章 重 点

- 算术平均数的计算
- 调和平均数的计算
- 标准差和标准差系数的计算

第一节 数据分布集中趋势的测度

所谓集中趋势是指一组数据向某一中心值靠拢的倾向，测度数据的集中趋势也就是寻找数据一般水

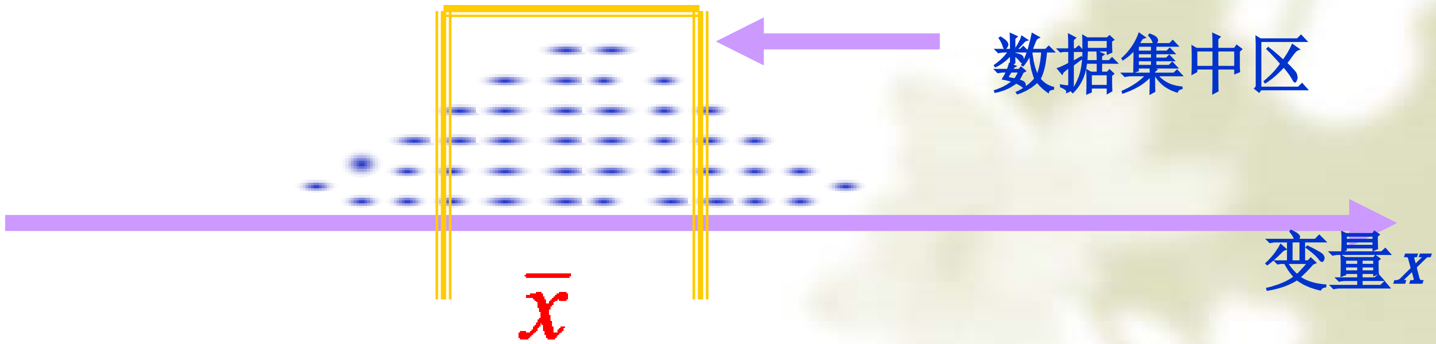
一、平均指标的意义和作用

1. 概念：平均指标是总体内各单位某一数量标志在具体时间、地点和条件下达到的一般水平的综合指标，又称平均数。

例：某班有10名学生，期末“统计学”成绩如下表：

学号	1	2	3	4	5	6	7	8	9	10
成绩	90	85	70	60	50	55	65	72	80	85

特点: 同质总体、差异抽象化、集中趋势性



2. 作用

- 利用平均指标便于进行对比分析
- 利用平均指标可以分析现象之间的依存关系
- 平均指标是制定定额的依据
- 利用平均指标可以进行数量上的推算

3. 种类

数值平均数
- 算术平均数 $\overline{X}$
- 调和平均数 $\overline{X}_h$
- 几何平均数 $\overline{X}_G$

位置平均数
- 众数 M_o
- 中位数 M_e

二、算术平均数

是总体标志总量除以总体单位总量所得到的平均数。

是计算平均指标最常用的方法和最基本的形式，是测度数据分布集中趋势应用最广的指标。

（一）计算公式

1. 简单算术平均数： $\bar{x} = \dfrac{\sum x}{n}$

适合于未分组资料

2.加权算术平均数：

例4-1-1：300户城市居民家庭拥有彩电的资料如下,试计算平均每户家庭的彩电数。

彩电数(台)	家庭数(户)
0	10
1	110
2	120
3	60
合　　计	300

解：$\overline{x} = \dfrac{0\times10+1\times110+2\times120+3\times60}{10+110+120+60}$

$$= \frac{530}{300} = 1.8(台/户)$$

$$\overline{x} = \frac{\sum xf}{\sum f}$$

例4-1-2：某企业职工按工资分组资料如下：

工资(元)	职工数(人)
400-500	50
500-600	70
600-700	120
700-800	60
合　　计	300

要求:计算全部职工的平均工资。

解: 计算过程如下:

工　资（元）	组中值 x	职工人数 f	x f
400—500	450	50	22500
500—600	550	70	38500
600—700	650	120	78000
700—800	750	60	45000
合　计	—	300	184000

$$\bar{x}=\frac{\sum xf}{\sum f}=\frac{184000}{300}=613.33(元)$$

（二）权数的选择

例4-1-3： 某管理局下属20个工业企业生产同一产品，其废品率的资料如下表：

废品率（%）	企业数(个)	产量(万件)
5以下	4	80
5-10	10	230
10-15	4	70
15以上	2	20

求：这20个企业的平均废品率

解：平均废品率 $= \dfrac{\text{废品总量}}{\text{总产量}} = \dfrac{\sum xf}{\sum f}$

$$= \frac{2.5\% \times 80 + 7.5\% \times 230 + 12.5\% \times 70 + 17.5\% \times 20}{400}$$

$= 7.875\%$

对绝对数求平均数时，频数一般都是权数。

但对相对数或平均数求平均数时，频数不一定是权数。

链接案例性别歧视

4. 例4-1-6：某管理局下属有27个企业，其销售与利润的情况如下：

利润率（%）	企业个数	销售额（万元）
8%以下	2	100
8%-10%	18	2000
10%-12%	6	900
12%以上	1	150

求这27个企业的平均利润率。

（三）影响加权算术平均数的因素

各组变量值（X）

各组权数所占比重 $\left(\frac{f}{\Sigma f}\right)$

与次数有没有关系？

(四)算术平均数的数学性质

1.算术平均数与总体单位数的乘积等于总体各单位标志值的总和。

简单算术平均数 $\bar{X}n = \sum X$ 加权算术平均数: $\bar{X}\sum f = \sum Xf$

2.如果每个变量值都加或减任意数值A，则平均数也要增多或减少这个数A。

简单算术平均数: $$\frac{\sum(X \pm A)}{n} = \frac{\sum X \pm nA}{n} = \bar{X} \pm A$$

加权算术平均数: $$\frac{\sum(X \pm A)f}{\sum f} = \frac{\sum Xf \pm A\sum f}{\sum f} = \bar{X} \pm A$$

3.如果每个变量值都乘以或除以一个任意值A,则平均数也乘以或除以这个数A。

简单算术平均数: $$\frac{\sum AX}{n} = \frac{A\sum X}{n} = A\bar{X} \qquad \frac{\sum\left(\frac{X}{A}\right)}{n} = \frac{\frac{1}{A}\sum X}{n} = \frac{\bar{X}}{A}$$

加权算术平均数: $$\frac{\sum AXf}{\sum f} = \frac{A\sum Xf}{\sum f} = A\bar{X} \qquad \frac{\sum\left(\frac{X}{A}\right)f}{\sum f} = \frac{\frac{1}{A}\sum Xf}{\sum f} = \frac{\bar{X}}{A}$$

4. 各个变量值与算术平均数的离差之和等于零

- 简单算术平均数:
- 证明:

$$\sum (X - \overline{X}) = 0$$

$$\sum(X-\overline{X})=\sum X-n\overline{X}=\sum X-n\cdot\frac{\sum X}{n}=\sum X-\sum X=0$$

- 加权算术平均数:
- 证明:

$$\sum (X - \overline{X})f = 0$$

$$\sum(X-\overline{X})f=\sum Xf-\overline{X}\sum f=\sum Xf-\frac{\sum Xf}{\sum f}\cdot\sum f=\sum Xf-\sum Xf=0$$

5. 各个变量值与算术平均数的离差平方之和等于最小值

- 简单算术平均数: $\sum (X-\overline{X})^2 = $ 最小值

  加权算术平均数: $\sum (X-\overline{X})^2 f = $ 最小值

- 证明: 设 X_0 为任意数, $X_0 = \overline{X} - C$, 则 $C = \overline{X} - X_0$

  以 X_0 为中心的离差平方之和为:

$$\sum (X - X_0)^2 = \sum \left[X - \left(\overline{X} - C\right)\right]^2$$

$$= \sum \left[\left(X - \overline{X}\right) + C\right]^2$$

- $$= \sum \left(X - \overline{X}\right)^2 + 2C\sum \left(X - \overline{X}\right) + nC^2$$

- $$= \sum \left(X - \overline{X}\right)^2 + nC^2$$

$$\because nC^2 \geq 0 \therefore \sum (X - X_0)^2 \geq \sum \left(X - \overline{X}\right)^2$$

$\sum \left(X - \overline{X}\right)^2$ 为最小值

同理, $\sum \left(X - \overline{X}\right)^2 f$ 为最小值

（五）算术平均数应用的特点

- 算术平均数适合代数方法的演算，不仅易于掌握，而且与大量的社会经济过程相适应。因此，应用十分广泛。
- 易受极端数值的影响，使算术平均数的代表性变小；而且受极大值的影响大于受极小值的影响。
- 当组距数列为开口组时，由于组中值不易确定，使平均数的代表性受影响。

三、调和平均数

（一）基本公式

例4-1-7：某蔬菜批发市场三种蔬菜的日成交数据如表，计算三种蔬菜该日的平均批发价格

某日三种蔬菜的批发成交数据			
蔬菜名称	批发价格(元) x_i	成交额(元) m_i	成交量(公斤) F_i
甲	1.20	18000	
乙	0.50	12500	
丙	0.80	6400	
合计	—	36900	

$$\bar{x}_H = \frac{\sum 成交额}{\sum \frac{成交额}{批发价格}}$$

$$= \frac{18000+12500+6400}{\frac{18000}{1.20}+\frac{12500}{0.50}+\frac{6400}{0.80}} = \frac{36900}{48000} = 0.77(元/斤)$$

得，加权调和平均数 $\bar{x}_H = \frac{\sum m}{\sum \frac{m}{x}}$

如m_i相等，得简单调和平均数 $\bar{x}_H = \frac{n}{\sum \frac{1}{x}}$

（二）调和平均数和算术平均数的判断

1. 例4-1-8：

过级率（%）	班数（个）	总人数	总过关人数
40以下	2	70	21
40-60	2	60	30
60-80	5	200	140
80-100	1	40	36

要求：(1) 舍弃总过关人数资料，求平均过级率；

(2) 舍弃总人数资料，求平均过级率。

解：平均过级率 $= \dfrac{\text{总过关人数}}{\text{总人数}}$

(1) $\bar{x} = \dfrac{0.3\times70+0.5\times60+0.7\times200+0.9\times40}{70+60+200+40} = \dfrac{227}{370} =$ 61.35%

即加权算术 $\dfrac{\sum xf}{\sum f}$

(2) $\bar{x} = \dfrac{21+30+140+36}{\dfrac{21}{0.3}+\dfrac{30}{0.5}+\dfrac{140}{0.7}+\dfrac{36}{0.9}} = \dfrac{227}{370} = 61.35\%$

即加权调和 $\dfrac{\sum m}{\sum \dfrac{m}{x}}$

2.结论：

已知x的文字公式中的分母（f）资料时，用加权算术平均数

已知x的文字公式中的分子（m）资料时，用加权调和平均数

调和平均数是算术平均数的变形

$$\bar{x} = \frac{\sum m}{\sum \frac{m}{x}}$$

$$\Longleftrightarrow \frac{\sum xf}{\sum \frac{xf}{x}} \Longleftrightarrow \frac{\sum xf}{\sum f}$$

(三)应用

- 1. 由相对数计算平均数时调和平均数法的应用：

例

某公司有四个工厂，已知其计划完成程度(%)及实际产值资料如下：

工厂	计划完成程度(%) X	实际产值(万元) m
甲	90	90
乙	100	200
丙	110	330
丁	120	480
合计	-	1,100

$$\text{平均完成计划程度}\ \frac{\sum m}{\sum \frac{1}{X} m} = \frac{1{,}100}{1{,}000} = 110\%$$

2. 由相对数计算平均数时加权平均数法的应用：

例 某公司有四个工厂，已知其计划完成程度(%)及计划产值资料如下：

工厂	计划完成程度 (%) X	计划产值 (万元) F
甲	90	100
乙	100	200
丙	110	300
丁	120	400
合计	-	1,000

$$\text{平均完成计划程度}\ \frac{\sum XF}{\sum F}=\frac{90\%\times100+100\%\times200+110\%\times300+120\%400}{1{,}000}=110\%$$

△ 调和平均数的特点

- 如果数列中有一标志值等于零，则无法计算$\overline{X}_h$；
- 较之算术平均数，$\overline{X}_h$受极端值的影响要小。

例如：1990年某月份甲、乙两农贸市场某农产品的价格、成交量和成交额资料如下：

品种	价格（元/千克）	甲市场成交额（万元）	乙市场成交量（千克）
甲	2.4	1.2	10 000
乙	2.8	2.8	5 000
丙	3	1.5	5 000
合计		5.5	20 000

试问哪一个市场农产品的平均价格较高？并说明原因

四、几何平均数

（一）概念

它是N个变量值的连乘积的N次方根。

常用于计算平均变化率或平均发展速度。

（二）计算公式

1. 未分组，用简单几何平均数 ：

$$\overline{X}_G = \sqrt[n]{X_1 \cdot X_2 \cdots X_n} = \sqrt[n]{\prod X}$$

2. 资料分组时，用加权几何平均数 ：

$$\overline{X}_G = \sqrt[\sum f]{X_1^{f_1} \cdot X_2^{f_2} \cdots X_n^{f_n}} = \sqrt[\sum f]{\prod x^f}$$

例1

某流水生产线有前后衔接的五道工序。某日各工序产品的合格率分别为95%、92%、90%、85%、80%，整个流水生产线产品的平均合格率为:

$$G=\sqrt[5]{0.95\times0.92\times0.90\times0.85\times0.80}$$

$$=\sqrt[5]{0.5349}=88.24\%$$

例2 加权几何平均数

投资银行某笔投资的年利率是按复利计算的，25年的年利率分配是：有1年为3%，有4年为5%，有8年为8%，有10年为10%，有2年为15%，求平均年利率。

年本利率(%) X	年数 f
103	1
105	4
108	8
110	10
115	2
合计	25

$$\bar{x}_g = \sqrt[25]{1.03 \; *1.05^4 *1.08^8 *1.1^{10} *1.15^2}$$

$=108.6\%$

这就是说，25年的平均本利率为108.6%，年平均利率即为8.6%。

△ 几何平均数的特点

- 如果数列中有一个标志值等于零或负值，就无法计算 $\bar{X}_G$；
- 受极端值的影响较 $\bar{X}$ 和 $\bar{X}_h$ 小；
- 它适用于反映特定现象的平均水平，即现象的总标志值是各单位标志值的连乘积。

五、切尾均值

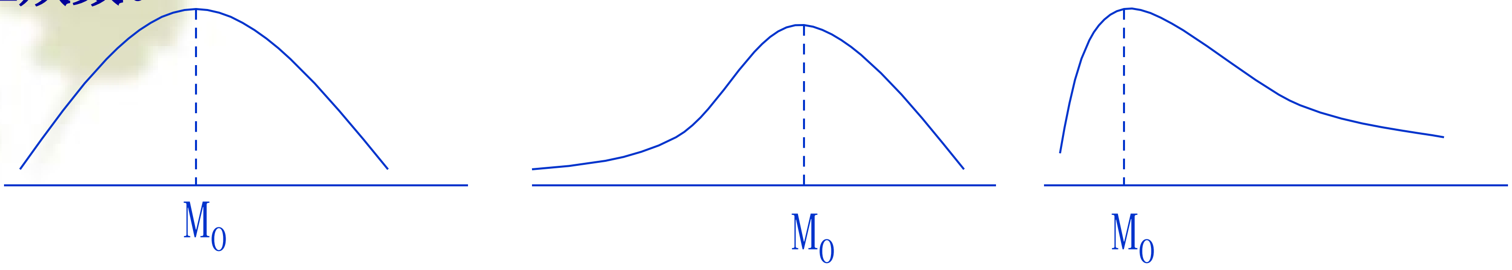
在日常生活中，我们经常遇到切尾均值这一平均数。切尾均值是切掉数据大小两端的若干数值后，就中间各项数值计算其算术平均数。这种平均数的测定方法在电视大奖赛、体育比赛及需要人们进行综合评价的竞赛项目中得到广泛应用。

六、众数　M_0

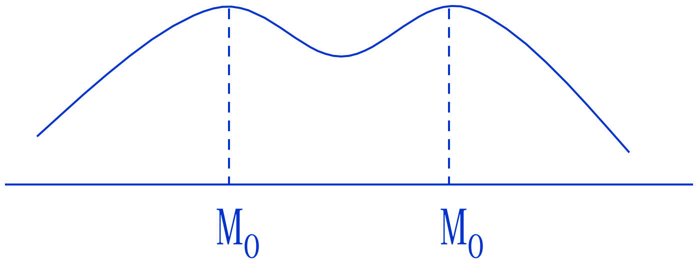
- 1. **概念**：众数是在总体中出现次数最多的那个标志值

由定义可看出众数存在的条件：

① 只有总体单位数比较多，而且又有明显的集中趋势时才存在众数。

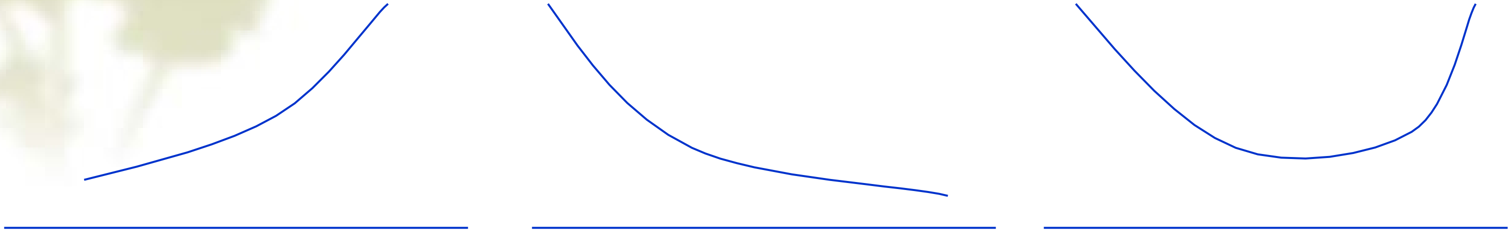


若有两个次数相等的众数，则称复众数。



② 在单位数很少，或单位数虽多但无明显集中趋势时，计算众数是没有意义的。

下三图无众数：



例:

X	f1	f2	f3
1	10	80	20
2	10	10	25
3	10	73	18
4	10	80	22
5	10	25	15