# 第四节　数据描述

# 第一章 第四节数据描述

| 年度 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|---|---|---|---|
| 考频 | 0 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |

# 第一章 第四节数据描述

一、平均值

二、极差、方差与标准差

三、数据的图表表示

# 一、平均值

1.中位数

将一组数据按照由小到大（或由大到小）的顺序排列，处于最中间位置的一个数据（或中间两个数据的平均数）.

2.众数

一组数据中出现次数最多的数据.

# 一、平均值

3.平均值（平均数）

统计学中反映一组数据的集中趋势或中心位置的度量.

平均值包含算术平均值、几何平均值和加权平均值等.

（1）算术平均值

设$n$个数$x_1$，$x_2$，…，$x_n$，称$\overline{x}=\frac{x_1+x_2+\cdots+x_n}{n}=\frac{\sum_{i=1}^{n}x_i}{n}$为这$n$个数的算术平均值.

①负数也存在算术平均值. ②0的算术平均值是0.

# 一.平均值

3.平均值（平均数）

（2）几何平均值

设$n$个数$x_1$，$x_2$，…，$x_n$，称$x_g = \sqrt[n]{x_1 x_2 \cdots x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$为这$n$个数的几何平均值.

根据定义可知：①负数不存在几何平均值.

②0不存在几何平均值.

# 一、平均值

3.平均值（平均数）

（3）加权平均值

设$n$个数$x_1$，$x_2$，…，$x_n$的权分别是$w_1$，$w_2$，…，$w_n$，

称$\overline{x}=\frac{x_1w_1+x_2w_2+\cdots+x_nw_n}{w_1+w_2+\cdots+w_n}$为这$n$个数的加权平均值.

其中，$w_1+w_2+\cdots+w_n=n$.

例：小明某科考试成绩：平时成绩80（占比20%），期中考试成绩90（占比30%），

期末考试成绩95（占比50%），则加权平均值为$\frac{80\times20\%+90\times30\%+95\times50\%}{20\%+30\%+50\%}=90.5$

## 二、极差、方差与标准差

1.极差

（1）定义：极差 = 最大值 - 最小值

（2）意义：用来反映一组数据变化范围的大小，只对极端值较为敏感.

2.方差

设$n$个数$x_1$，$x_2$，…，$x_n$，则方差为：

$$S^2=\frac{1}{n}[(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2].$$

**先平均，再求差，然后平方，最后再平均.**

# 二、极差、方差与标准差

2.方差

（1）扩展公式

$$S^2=\frac{1}{n}[(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2]$$

$$=\frac{1}{n}[{x_1}^2+{x_2}^2+\cdots+{x_n}^2-n(\overline{x})^2]$$

$$=\frac{{x_1}^2+{x_2}^2+\cdots+{x_n}^2}{n}-\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)^2$$

# 二、极差、方差与标准差

2.方差

（2）平均数与方差的性质

原数据：$x_1$，$x_2$，…，$x_n$的平均数为$\bar{x}$，方差为$S^2$

①新数据：$ax_1$，$ax_2$，…，$ax_n$的平均数是$a\bar{x}$，方差为$a^2S^2$.

②新数据：$x_1+b$，$x_2+b$，…，$x_n+b$的平均数为$\bar{x}+b$，方差为$S^2$.

③新数据：$ax_1+b$，$ax_2+b$，…，$ax_n+b$的平均数为$a\bar{x}+b$，方差为$a^2S^2$.

# 二、极差、方差与标准差

2.方差

（2）平均数与方差的性质

原数据：$x_1$，$x_2$，…，$x_n$的平均数为$\bar{x}$，方差为$S^2$

③新数据：$ax_1+b$，$ax_2+b$，…，$ax_n+b$的平均数为$a\bar{x}+b$，方差为$a^2S^2$.

# 二、极差、方差与标准差

2.方差

（2）平均数与方差的性质

原数据：$x_1$，$x_2$，…，$x_n$的平均数为$\bar{x}$，方差为$S^2$

③新数据：$ax_1+b$，$ax_2+b$，…，$ax_n+b$的平均数为$a\bar{x}+b$，方差为$a^2S^2$.

# 二、极差、方差与标准差

3.标准差

$$S=\sqrt{S^2}=\sqrt{\frac{1}{n}[(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2]}$$

$$=\sqrt{\frac{{x_1}^2+{x_2}^2+\cdots+{x_n}^2}{n}-\overline{x}^2}$$

# 二、极差、方差与标准差

4.方差与标准差的异同

（1）相同之处：都是反应一组数据离散程度的统计量.

（2）不同之处：单位不一致

方差的单位是数据单位的平方，标准差的单位与所研究数据单位一致.

# 二、极差、方差与标准差



| | |
|---|---|
| **平均数** | 反映数据的平均水平 |
| **众数** | 反映数据的集中趋势 |
| **中位数** | 反映数据的中间值 |
| **方差** | 反映数据的波动大小，方差**大**，波动**大**；方差**小**，波动**小** |
| **标准差** | 反映数据的波动大小，标准差**大**，波动**大**；标准差**小**，波动**小** |
| 平均数、众数和中位数：反映数据的总体趋势 | |

## 二.极差.方差与标准差

【例1】甲、乙、丙三人每轮各投篮10次，投了三轮.投中数如下表：

| | 第一轮 | 第二轮 | 第三轮 |
|---|---|---|---|
| 甲 | 2 | 5 | 8 |
| 乙 | 5 | 2 | 5 |
| 丙 | 8 | 4 | 9 |

记$\sigma_1$，$\sigma_2$，$\sigma_3$分别为甲、乙、丙投中数的方差，则（　　）

A.$\sigma_1>\sigma_2>\sigma_3$　　B.$\sigma_1>\sigma_3>\sigma_2$　　C.$\sigma_2>\sigma_1>\sigma_3$

D.$\sigma_2>\sigma_3>\sigma_1$　　E.$\sigma_3>\sigma_2>\sigma_1$

# 三.数据的图表表示

1.数据定义

（1）频数：是指某个数据出现的次数.

（2）频率：频数与数据总个数之比.

2.数据的图表表示

（1）直方图

（2）饼图

（3）柱状图

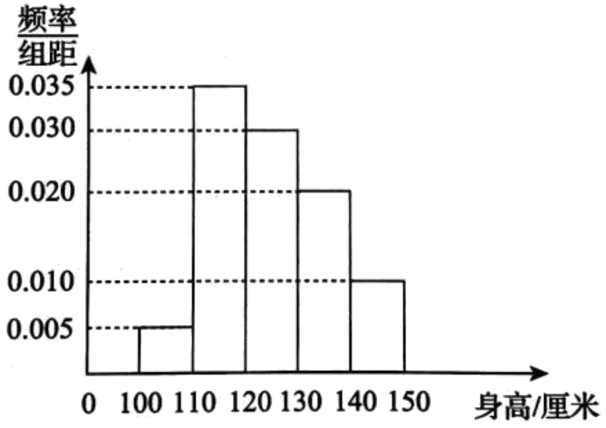
（4）数表

# 三.数据的图表表示

## （1）直方图

把数据分成若干个小组，每组的组距保持一致，并在直角坐标系的横轴上标出每组的位置（以组距作为底），计算每组所包含的数据个数（频数），以该组的“频率/组距”为高作矩形，这样得出若干个矩形构成的图叫做直方图.

- ✓ 组距的确定：一般是人为确定，不能太大也不能太小.
- ✓ 组数的确定：组数＝极差/组距.
- ✓ 每组频率的确定：频率＝频数/数据容量.
- ✓ 每组所确定的矩形的面积＝组距×$\frac{频率}{组距}$＝频率.
- ✓ 频率直方图下的总面积等于1（各个矩形面积之和等于1）.
- ✓ 分组时要遵循“不重不漏”的原则.



# 三.数据的图表表示

（1）直方图

- 众数是最高矩形底边中点的横坐标
- 中位数左边和右边的直方图的面积相等
- 平均数是直方图的重心，它等于每个小矩形的面积乘以小矩形底边中点横坐标之和.

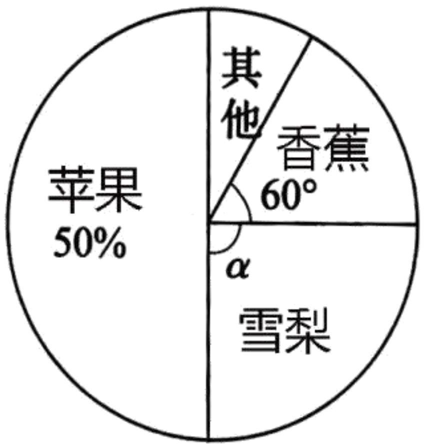
# 三.数据的图表表示

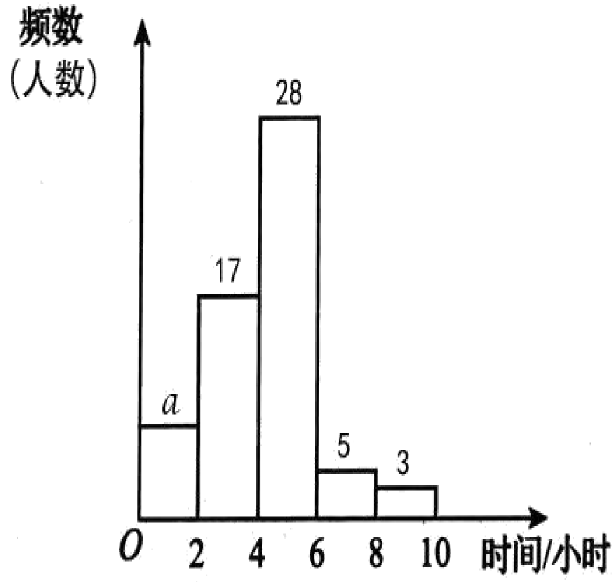
（2）饼图

饼图能显示部分在总体中所占百分比.

（3）柱状图

柱状图能显示每组中的具体数据.



# 三.数据的图表表示

## （4）数表

数表能显示数据的频率.

| | | |
|---|---|---|
| 销售额/万元 | 3 | 4 |
| 销售员人数 | 1 | 3 |

# 第一章 算术

- 第一节 实数
- 第二节 分数、小数、百分数
- 第三节 比与比例
- 第四节 数据描述
  - 一、平均值
  - 二、极差、方差与标准差
  - 三、数据的图表表示
- 第五节 数轴与绝对值