平均差与标准差

授课教师：　洪兴建

浙江财经大学数据科学学院

引例

稳定性

历史上我国奥运首金运动员是谁？

季军是谁？

六届奥运会2金3银1铜

为什么射击常见爸妈级选手？

一方面成绩好

另一方面稳定性高

离散指标

含义与作用

含义	➢ 反映数值**差异程度**的指标，亦称变异指标。
作用	➢ 衡量平均水平的**代表性** ➢ 反映均衡性和**稳定性** ➢ 为**统计推断**提高依据

全距

1 全距

- 全距（Range）或极差

$$R = 最大值 - 最小值 = Max\text{-}Min$$

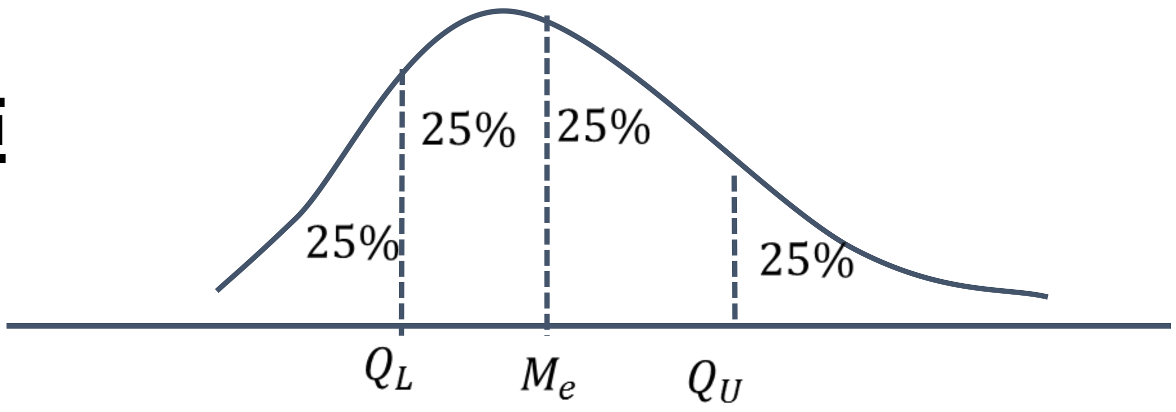
- **优点：**简单明了
- **缺点：**只反映变动幅度，最易受极端值影响

四分位差

2 四分位差

- 四分位差也称为内距

$$Q_d = Q_U - Q_L$$



- 特点

中间50%个体的变动幅度，不受极值影响。

异众比率

3 异众比率

- 非众数组频数占总频数之比：

$$V_r = \frac{\sum f_i - f_{mo}}{\sum f_i} = 1 - \frac{f_{mo}}{\sum f_i}$$

- 反映众数的代表性

平均差

4 平均差

- 即平均绝对偏差（Mean Absolute Deviation）

在资料未分组的情况下，计算公式

$$MAD = \frac{\sum \left| x_i - \bar{x} \right|}{n}$$

在资料已分组的情况下，计算公式

$$MAD = \frac{\sum \left| x_i - \bar{x} \right| \cdot f_i}{\sum f_i}$$

离散指标

平均差

例1　第一组：600，700，800，900，1000

　　　第二组：630，670，800，900，1000

平均消费都是800，哪一组差异较小？

$$\text{第一组MAD}=\frac{|600-800|+|700-800|+|800-800|+|900-800|+|1000-800|}{5}=120\text{（元）}$$

$$\text{第二组MAD}=\frac{|630-800|+|670-800|+|800-800|+|810-800|+|820-800|}{5}=120\text{(元)}$$

平均差

平均差特点

（1）反映全部标志值变动

（2）受水平高低、计量单位影响

（3）性能没有标准差好

方差标准差

5 计算公式

- 方差即离差平方的算术平均数，标准差是方差的算术平方根。

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

or

$$\sigma^2=\frac{\sum_{i=1}^{K}(X_i-\bar{X})^2 f_i}{\sum_{i=1}^{K} f_i}$$

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

or

$$\sigma=\sqrt{\frac{\sum_{i=1}^{K}(X_i-\bar{X})^2 f_i}{\sum_{i=1}^{K} f_i}}$$

方差标准差

样本方差和样本标准差

$$s_{n-1}^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$$

or

$$s_{n-1}^2=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 f_i}{\sum_{i=1}^{k} f_i-1}$$

$$s_{n-1}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$$

or

$$s_{n-1}=\sqrt{\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 f_i}{\sum_{i=1}^{k} f_i-1}}$$

方差标准差

例2

谁的发挥更为稳定

序号	姚明	格伦-戴维斯	克里斯-保罗
	得分	得分	得分
1	19	7	12
2	12	13	4
3	28	12	32
4	17	15	14
5	15	23	21
6	21	21	26
7	7	10	9
8	11	14	31
9	24	26	42
10	23	18	29

方差标准差

$$\bar{x}_{姚}=\frac{\sum x_i}{n}=\frac{19+12+\cdots+23}{10}=17.7(分)$$

$$s_{姚}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{(19-17.7)^2+(12-17.7)^2+\cdots(23-17.7)^2}{10-1}}=6.55(分)$$

$$\bar{x}_{戴}=\frac{\sum x_i}{n}=\frac{7+13+\cdots+18}{10}=15.9(分)$$

$$s_{戴}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{(7-15.9)^2+(13-15.9)^2+\cdots(18-15.9)^2}{10-1}}=6.01(分)$$

$$\bar{x}_{保}=\frac{\sum x_i}{n}=\frac{12+4+\cdots+29}{10}=22(分)$$

$$s_{保}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}=\sqrt{\frac{(12-22)^2+(4-22)^2+\cdots(29-22)^2}{10-1}}=12.04(分)$$

✓从标准差看，戴维斯发挥更为稳定！

方差标准差

简便公式

$$S^2=\overline{x^2}-\bar{x}^2$$

其中 $\overline{x^2}=\dfrac{\sum_{i=1}^{n}x_i^2}{n}$ or $\dfrac{\sum_{i=1}^{n}x_i^2 f_i}{\sum_{i=1}^{n}f_i}$

推导

$$S^2=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\frac{\sum_{i=1}^{n}(x_i^2-2x_i\bar{x}+\bar{x}^2)}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}-\bar{x}^2$$

方差标准差

● **性质**

（1）常数的方差为0： $s_a^2=0$

$$S_x^2=\frac{\sum_{i=1}^{n}（x_i-\bar{x}）^2}{n}$$

（2）若$y=a+bx$，a,b**为常数，则**

$$s_y^2=b^2s_x^2$$

（3）标准差是计算标准化值的依据

$$z_i=\frac{x_i-\bar{x}}{s} \Longrightarrow \begin{array}{l}\bar{z}=0\\ s_z^2=1\end{array}$$

离散指标

方差标准差

优点

- 反应灵敏
- 适合代数计算
- 能够推断

缺点

- 受量纲约束
- 受水平影响

讨论

平均差与标准差

- 哪个好？ 很多人说标准差
- 为什么？ ~~计算方便~~ ~~因为"标准"~~ ⇒判断准则？

◆ 1原始：600，700，800，900，1000

◆ 2转移：630，670，800，900，1000

$$\boxed{MAD_1 = MAD_2 = 120}$$

*PD*转移公理

$$\boxed{\sigma_1 = 141.42 > \sigma_2 = 138.42}$$