平均数、中位数和众数的选用

1. 众数、样本数据的频率分布直方图最高的矩形的中点

2.中位数：在频率分布直方图中，中位数左边和右边的直方图的面积应该相等，由此可以估计中位数的值.

3.平均数：等于频率分布直方图中每个小矩形的面积乘以小矩形底边中点的横坐标之和。

任何一个样本数据的改变都会引起平均数的改变.

4.标准差计算公式：标准差是样本数据到平均数的一种平均距离，一般用 s 表示.

$$s=\sqrt{\frac{1}{n}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\right]}.$$

例 1 画出下列四组样本数据的直方图，说明它们的异同点.

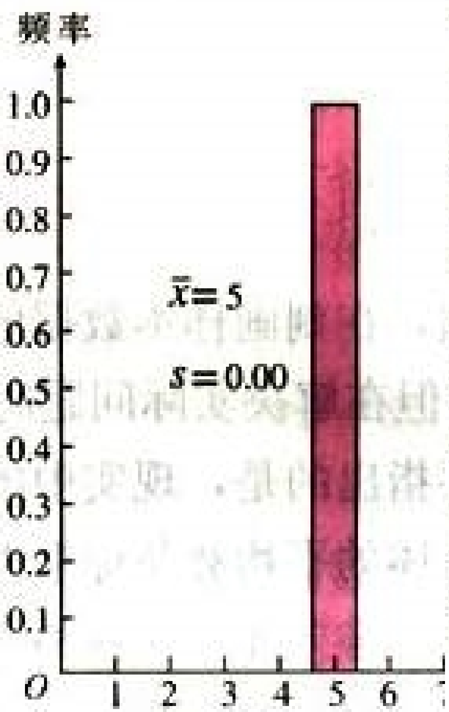
(1) 5，5，5，5，5，5，5，5，5；

(2) 4，4，4，5，5，5，6，6，6；

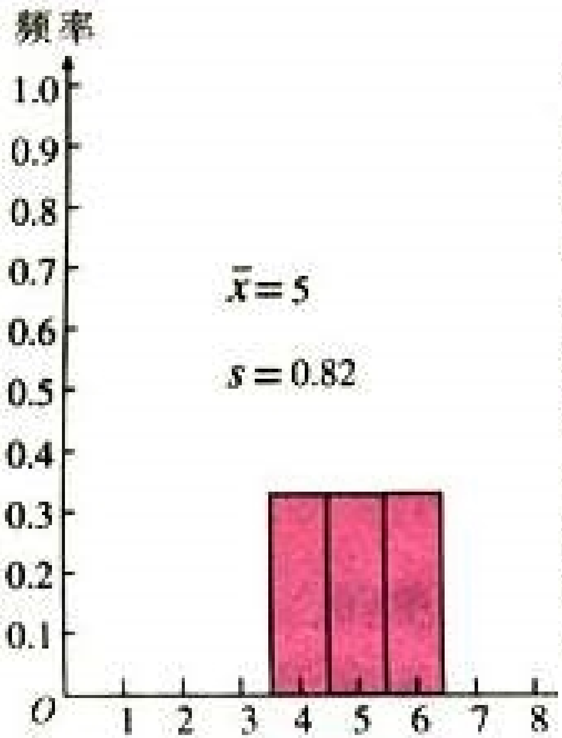
(3) 3，3，4，4，5，6，6，7，7；

(4) 2，2，2，2，5，8，8，8，8.

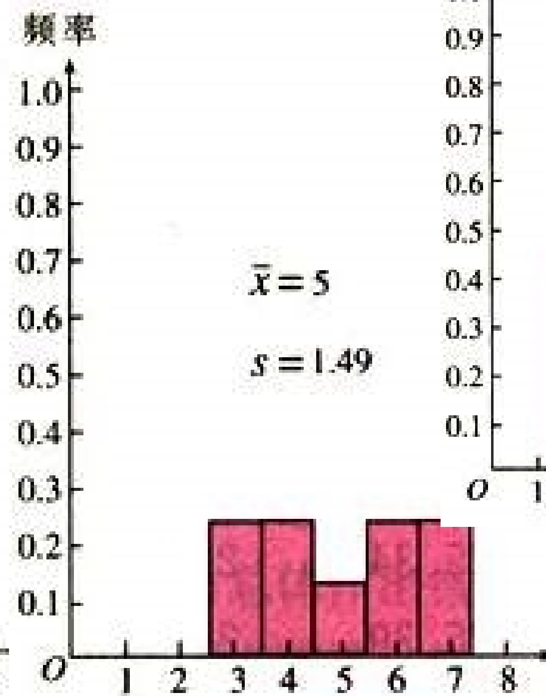
解：四组样本数据的直方图是：



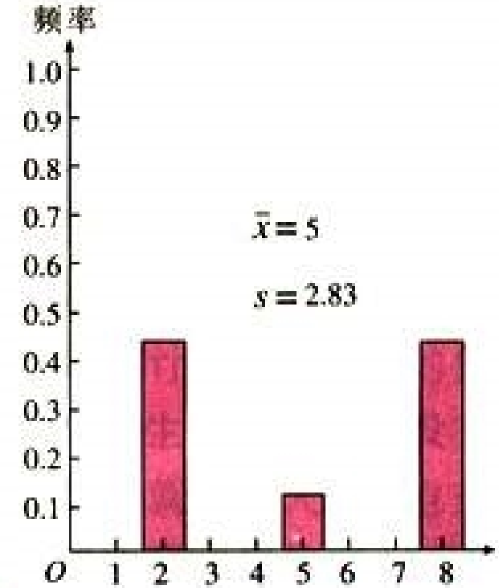
(1)



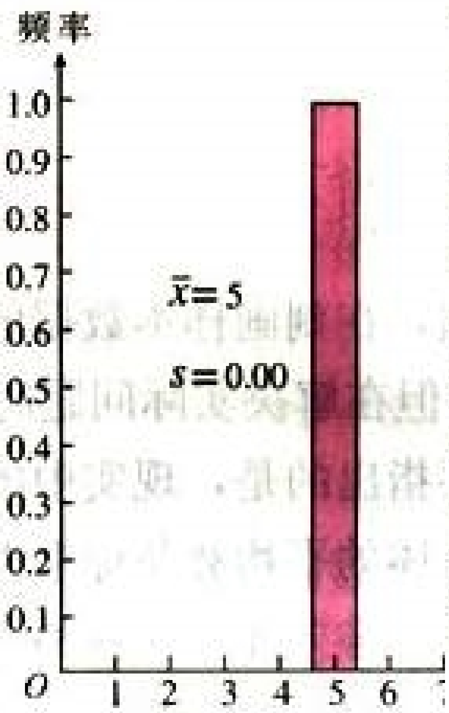
(2)



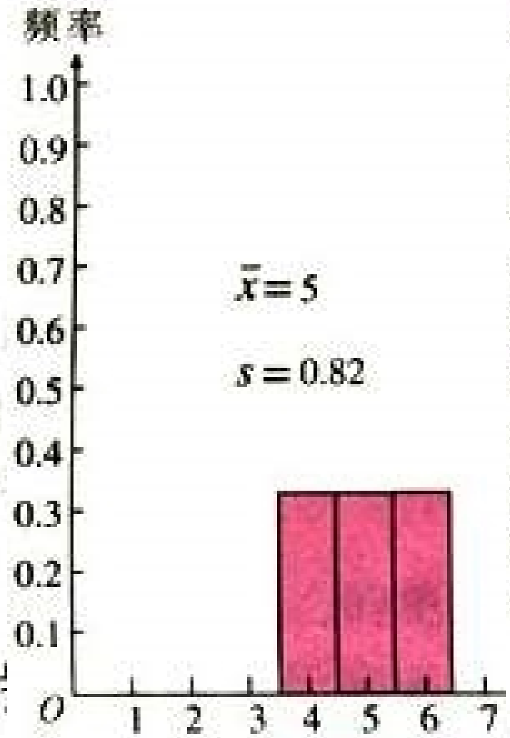
(3)



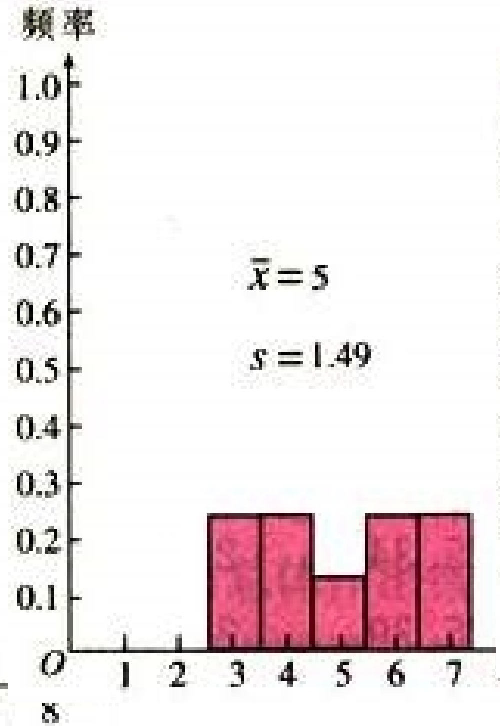
(4)



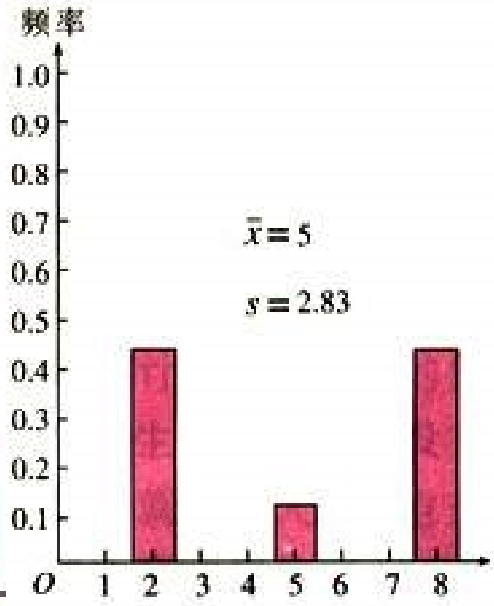
(1)



(2)



(3)



(4)

四组数据的平均数都是 5.0，标准差分别是 0.00，0.82，1.49，2.83. 虽然它们有相同的平均数，但是它们有不同的标准差，说明数据的分散程度是不一样的.

标准差还可以用于对样本数据的另外一种解释. 例如，在关于居民月均用水量的例子中，平均数 $\bar{x}=1.973$，标准差 $s=0.868$，所以

$$\bar{x}+s=2.841,\quad \bar{x}+2s=3.709;$$

$$\bar{x}-s=1.105,\quad \bar{x}-2s=0.237.$$

这 100 个数据中，在区间 $[\bar{x}-2s, \bar{x}+2s]=[0.237, 3.709]$ 外的只有 4 个，也就是说，$[\bar{x}-2s, \bar{x}+2s]$ 几乎包含了所有样本数据.

从数学的角度考虑，人们有时用标准差的平方 s^2——方差来代替标准差，作为测量样本数据分散程度的工具：

$$s^2=\frac{1}{n}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\right].$$

通常的做法是用样本的平均数和标准差去估计总体的平均数与标准差.

例 2　甲乙两人同时生产内径为 25.40 mm 的一种零件. 为了对两人的生产质量进行评比，从他们生产的零件中各抽出 20 件，量得其内径尺寸如下（单位：mm）：

甲

25.46	25.32	25.45	25.39	25.36
25.34	25.42	25.45	25.38	25.42
25.39	25.43	25.39	25.40	25.44
25.40	25.42	25.35	25.41	25.39

乙

25.40	25.43	25.44	25.48	25.48
25.47	25.49	25.49	25.36	25.34
25.33	25.43	25.43	25.32	25.47
25.31	25.32	25.32	25.32	25.48

从生产的零件内径的尺寸看，谁生产的质量较高？

解：用计算器计算可得

$$\bar{x}_{甲}=25.400\,5,\qquad \bar{x}_{乙}=25.400\,8;$$

$$s_{甲}=0.038,\qquad s_{乙}=0.074.$$

从样本平均数看，甲生产的零件内径比乙的更接近内径标准(25.40 mm)，但是差异很小；从样本标准差看，由于$s_{甲}<s_{乙}$，因此甲生产的零件内径比乙的稳定程度高得多．于是，可以作出判断，甲生产的零件的质量比乙的高一些．

练 习

1. 农场种植的甲乙两种水稻，在面积相等的两块稻田中连续6年的年平均产量如下(单位：500 g)：

品种	第1年	第2年	第3年	第4年	第5年	第6年
甲	900	920	900	850	910	920
乙	890	960	950	850	860	890

哪种水稻的产量比较稳定？

解：甲乙两种水稻6年平均产量的平均数都是900，但甲的标准差约等于23.8，乙的标准差约等于41.6，所以甲的产量比较稳定。

2. 一个小商店从一家食品有限公司购进21袋白糖，每袋白糖的标准重量是500 g，为了了解这些白糖的重量情况，称出各袋白糖的重量（单位：g）如下：

486	495	496	498	499	493	493
498	484	497	504	489	495	503
499	503	509	498	487	500	508

求：

（1）21袋白糖的平均重量 $\bar{x}$ 是多少？标准差 s 是多少？

（2）重量位于 $\bar{x}-s$ 与 $\bar{x}+s$ 之间有多少袋白糖？所占的百分比是多少？

解:(1)平均重量 $\bar{x} \approx 496.86$ **,**标准差**s≈6.55.**

(2)重量位于$(\bar{x}-s,\bar{x}+s)$之间有**14**袋白糖，所占的百分比约为**66.67%.**

3. 下列数据是30个不同国家中每100 000名男性患某种疾病的死亡率：

27.0	23.9	41.6	33.1	40.6	18.8	13.7	28.9	13.2	14.5
27.0	34.8	28.9	3.2	50.1	5.6	8.7	15.2	7.1	5.2
16.5	13.8	19.2	11.2	15.7	10.0	5.6	1.5	33.8	9.2

(1) 作出这些数据分布的频率分布直方图；

(2) 请由这些数据计算平均数、中位数和标准差，并对它们的含义进行解释.

(1)略

(2)平均数$\overline{x} \approx 19.25$，中位数为**15.2**，标准差**s≈12.50.**

这些数据表明这些国家男性患该病的平均死亡率约为**19.25**，有一半国家的死亡率为不超过**15.2**，$\overline{x} > 15.2$说明存在大的异常数据，值得关注。这些异常数据使标准差增大。