2.2.2　用样本的数字特性预计总体的数字特性

（一）众数、中位数、平均数

一 众数、中位数、平均数的概念

众数：在一组数据中，出现次数最多的数据叫做这组数据的众数.

中位数：将一组数据按大小依次排列，把处在最中间位置的一种数据（或最中间两个数据的平均数）叫做这组数据的中位数.

平均数：一组数据的算术平均数,即

$$\overline{X}=\frac{1}{n}(x_1+x_2+\cdots+x_n)$$

练习

1.甲在一次射击比赛中的得分以下:(单位:环).7, 8, 6, 8, 6, 5, 9, 10, 7, 5,则他命中的平均数是___7.1___,中位数是___7___

2. 众数是___5, 6, 7, 8___

2. 某次数学试卷得分抽样中得到:90分的有3个人,80分的有10人,70分的有5人,60分的有2人,则这次抽样的平均分为77分

_______.

问题1：众数、中位数、平均数这三个数普通都会来自于同一种总体或样本，它们能表明总体或样本的什么性质？

众数:反映的往往是局部较集中的数据信息

中位数:是位置型数，反映处在中间部位的数据信息

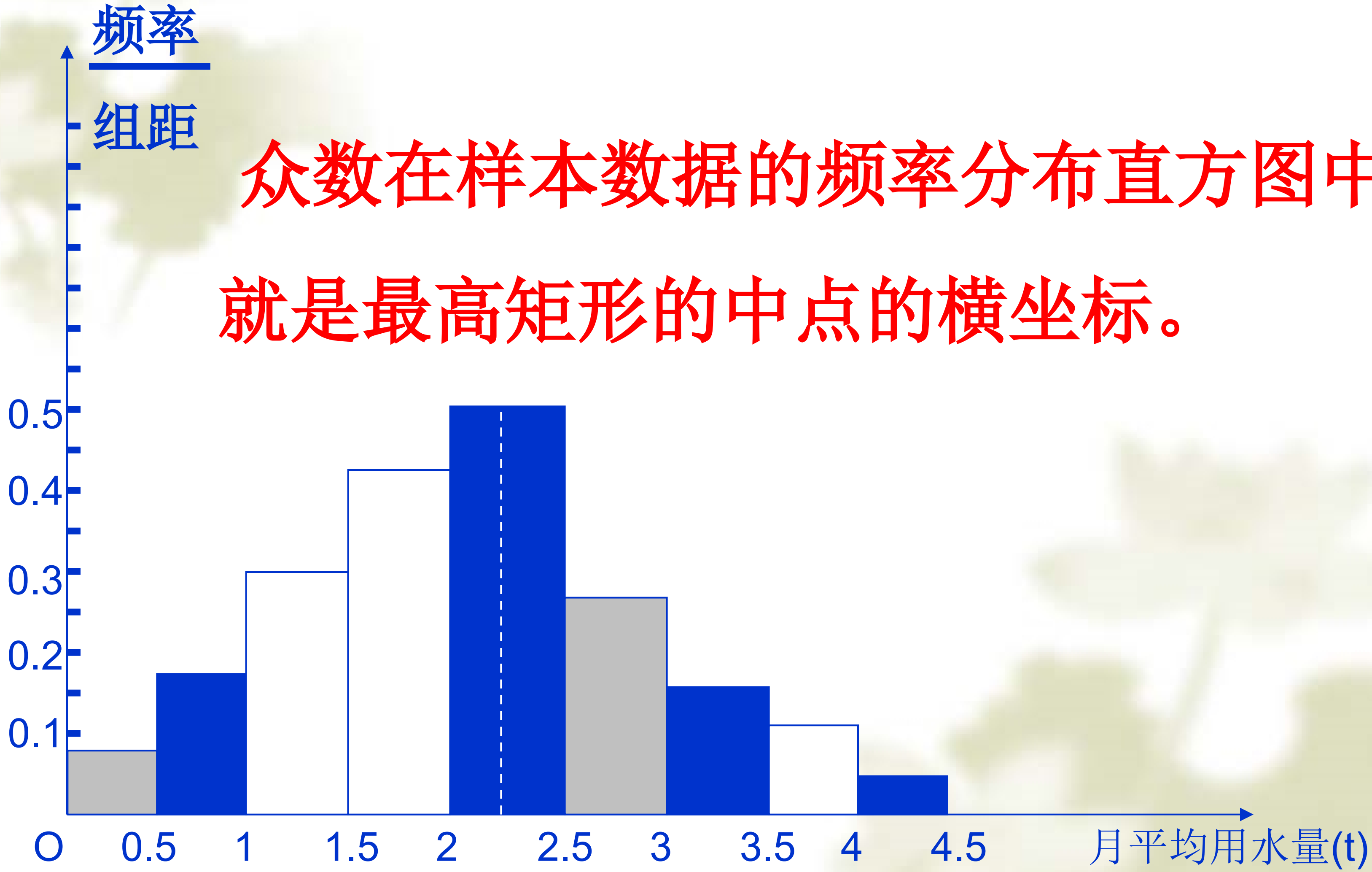
平均数:反映全部数据的平均水平

二、众数、中位数、平均数与频率分布直方图的关系

如何在频率分布直方图中预计众数

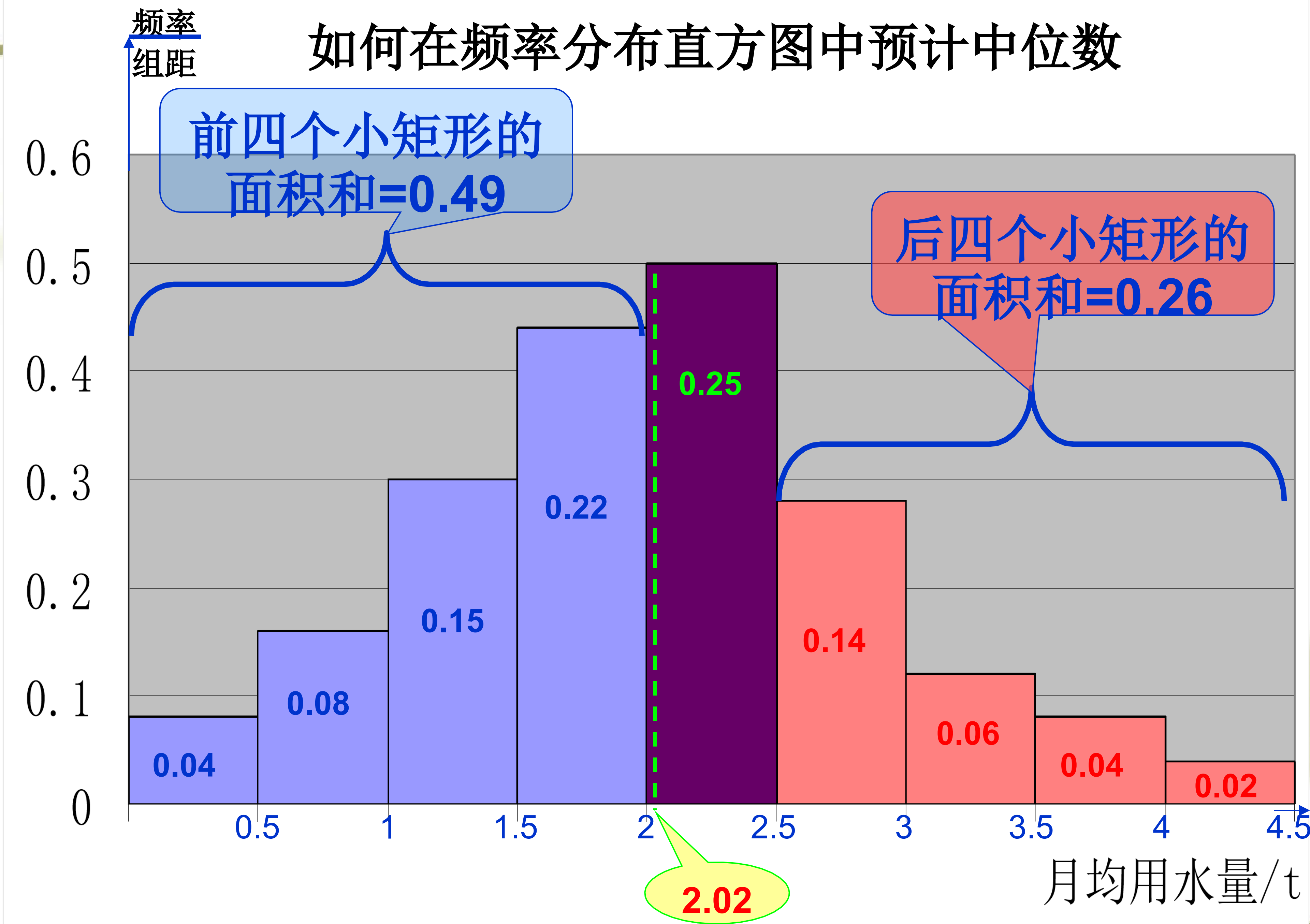
众数在样本数据的频率分布直方图中，

就是最高矩形的中点的横坐标。



可将众数看作直方图中面积最大长方形的“中心”

如何在频率分布直方图中预计中位数



分组	频率
[0, 0.5)	0.04
[0.5, 1)	0.08
[1, 1.5)	0.15
[1.5, 2)	0.22
[2, 2.5)	0.25
[2.5, 3)	0.14
[3, 3.5)	0.06
[3.5,	0.04
	0.02
	1

在样本中中位数的左右各有50%的样本数，条形面积各为0.5,因此反映在直方图中位数左右的面积相等.

$$0.04+0.08+0.15+0.22=0.49$$

$$x=0.02$$

中位数 $=2+0.02=2.02$

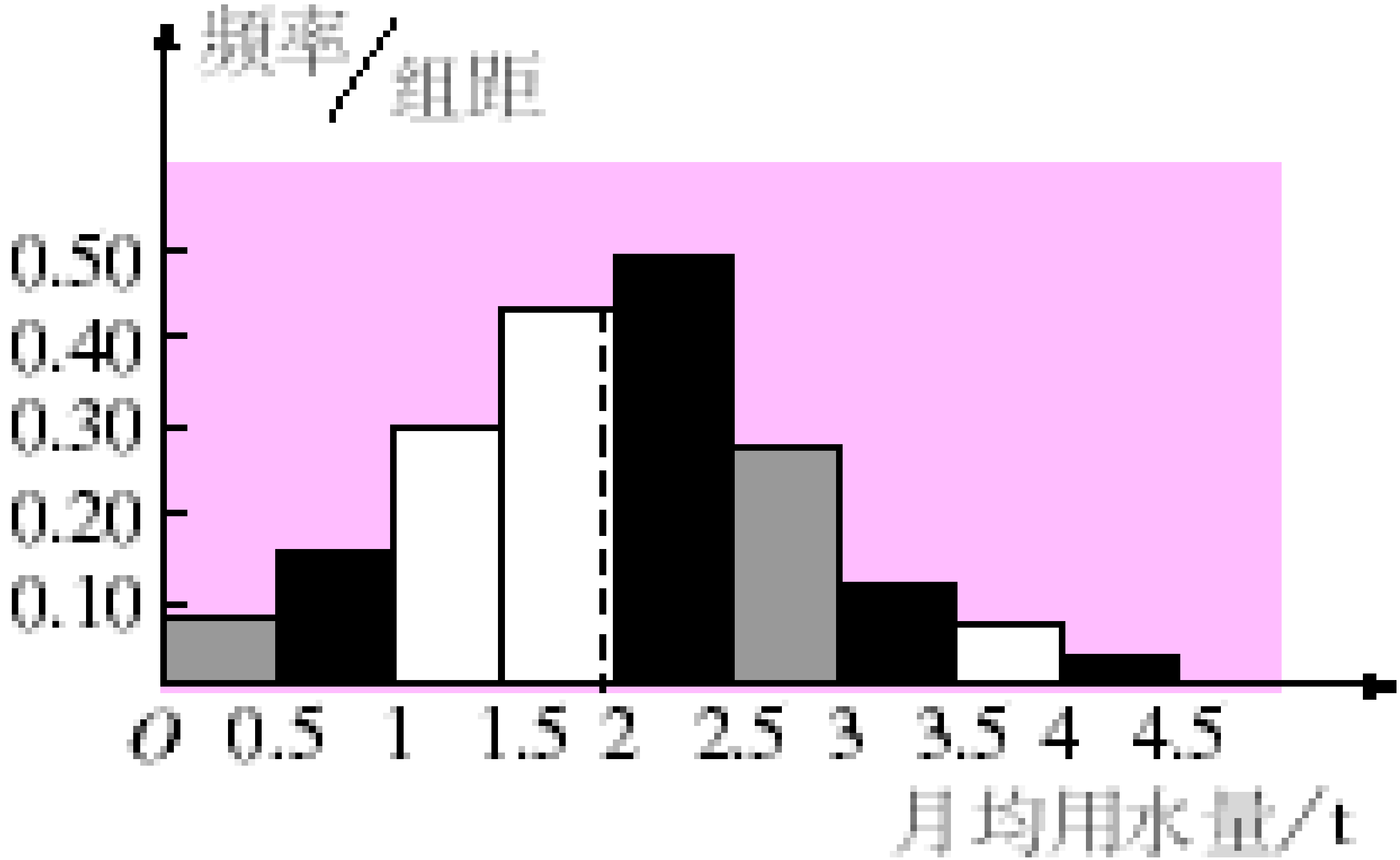
可将中位数看作整个直方图面积的“中心”

思考讨论下列问题：

1、2.02这个中位数的预计值，与样本的中位数值2.0不同，你能解释其中因素吗？

答：2.02这个中位数的预计值,与样本的中位数值2.0不同，这是由于样本数据的频率分布直方图，只是直观地表明分布的形状，但是从直方图本身得不出原始的数据内容，直方图已经损失某些样本信息。因此由频率分布直方图得到的中位数预计值往往与样本的实际中位数值不一致.

如何在频率分布直方图中预计平均数



分组	频数累计	频数	频率
[0, 0.5)	正 (4)	4	0.04
[0.5, 1)	正下	8	0.08
[1, 1.5)	正正正	15	0.15
[1.5, 2)	正正正正丅	22	0.22
[2, 2.5)	正正正正正	25	0.25
[2.5, 3)	正正𠁥	14	0.14
[3, 3.5)	正一	6	0.06
[3.5, 4)	𠁥	4	0.04
[4, 4.5]	丅	2	0.02
合计		100	1.00

$$\bar{x}=\frac{1}{100}(x_1+x_2+\cdots+x_{100})=\frac{1}{100}\left[(x_1+\cdots+x_4)+(x_5+\cdots+x_{12})+\cdots+(x_{99}+x_{100})\right]$$

$$=\frac{4}{100}\bar{x}_{1-4}+\frac{8}{100}\bar{x}_{5-12}+\cdots+\frac{2}{100}\bar{x}_{99-100}$$

平均数的预计值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。

$$=0.04\,\frac{0+0.5}{2}+0.08\,\frac{0.5+1}{2}+\cdots+0.02\,\frac{4+4.5}{2}$$

=2.02

可将平均数看作整个直方图面积的“重心”

例1：从甲、乙、丙三个厂家的同一种产品抽取8件，对其使用寿命进行追踪调查，成果以下：

甲：3,4,5,6,8，8,8,10

乙：4,6,6,6,8,9,12,13

丙：3,3,4,7,9,10,11,12

三个厂家广告中均称该产品的使用寿命为8年，请根据成果判断厂家在广告中分别运用了那些特性数？

若不考虑其它因素，你会选择哪个厂家的产品，说出理由。

思考讨论下列问题：

2、样本中位数不受少数极端值的影响，这在某些状况下是一种优点，但它对极端值的不敏感有时也会成为缺点。你能举例阐明吗？

答：优点：对极端数据不敏感的办法能够有效地防止错误数据的影响。

对极端值不敏感有利的例子:例如当样本数据质量比较差，即存在某些错误数据（如数据录入错误、测量错误等）时，用抗极端数据强的中位数表达数据的中心值更精确

缺点：（1）出现错误的数据也不懂得；2）对极端值不敏感有弊的例子：某人含有初级计算机专业技术水平，想找一份收入好的工作。这时如果采用各个公司计算机专业技术人员收入的中位数作为选择工作的参考指标就会冒这样的风险：

很可能所选择公司的初级计算机专业技术水平人员的收入很低，其因素是中位数对极小的数据不敏感。这里更加好的办法是同时用平均工资和中位数作为参考指标，选择平均工资较高且中位数较大的公司就业.

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/517120026051006163