第29章 数据的分析

29.1 数据的代表

知识清单

知识1 平均数
知识2 加权平均数
知识3 中位数
知识4 众数
知识5 平均数、中位数、众数的优缺点

知识1 平均数

一般地，对于 n 个数 $x_1,x_2,\cdots,x_n$，我们把$\frac{1}{n}(x_1+x_2+\cdots+x_n)$叫做这 n 个数的算术平均数，简称平均数，记作“$\bar{x}$”，读作“x 拔”.

温馨提示

①平均数、数的个数以及所有数的总和这三个量中，已知任意两个就能求出第三个，平均数$=\frac{\text{所有数的总和}}{\text{数的个数}}$.

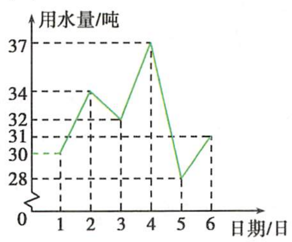
②平均数是描述一组数据的一种常用指标.一组数据的平均数只有一个.

③平均数的大小与一组数据里的每个数据均有关系，其中任一数据的变动都会引起平均数的变动.平均数容易受个别极端值影响.

④数据 $x_1,x_2,\cdots,x_n$ 的平均数为$\bar{x}$，则 $x_1\pm a,x_2\pm a,\cdots,x_n\pm a$ 的平均数为$\bar{x}\pm a$；$kx_1,kx_2,\cdots,kx_n$ 的平均数为$k\bar{x}$(a,k 为常数).

⑤总体中所有个体的平均数叫做总体平均数，样本中所有个体的平均数叫做样本平均数，通常用样本平均数去估计总体平均数.

例1 某住宅小区六月份中 1 至 6 日每天用水量变化情况如折线图所示，那么这 6 天的平均每天用水量是（　　）



A.30 吨　　B.31 吨
C.32 吨　　D.33 吨

解析 根据平均数的定义可得这 6 天平均每天的用水量为$\frac{30+34+32+37+28+31}{6}=32$(吨)，故选 C.

答案 C

知识2 加权平均数

当一组数据中有数据重复出现时，如在 n 个数据中，x_1 出现f_1 次，x_2 出现f_2 次，……，x_k 出现 f_k 次(这里 $f_1+f_2+\cdots+f_k=n$)，那么这 n 个数据的平均数可表示为$\frac{x_1f_1+x_2f_2+\cdots+x_kf_k}{n}$，这个平均数也叫做加权平均数，其中$f_1$，$f_2$，$\cdots$，$f_k$ 分别叫做 $x_1,x_2,\cdots,x_k$ 的权.

或者，若 n 个数 $x_1,x_2,\cdots,x_n$ 的权分别是 $w_1,w_2,\cdots,w_n$，则$\frac{x_1w_1+x_2w_2+\cdots+x_nw_n}{w_1+w_2+\cdots+w_n}$叫做这 n 个数的加权平均数.

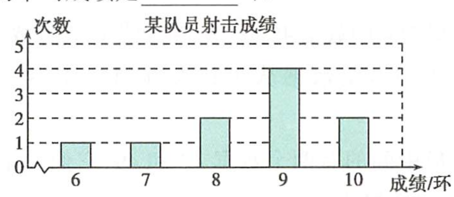
温馨提示

①加权平均数实际上是算术平均数的另一种表现形式.

②若各个数据的权相同，则加权平均数就是算术平均数，因而可以看出算术平均数实质上是加权平均数的一种特例.

③算术平均数是用一组数据的和除以数据的个数来计算的；加权平均数在计算上与算术平均数有所不同，是因为在实际问题中数据的“重要程度”未必相同，即各个数据的“权”未必相同.

例2 射击比赛中，某队员 10 次射击成绩如图所示，则该队员的平均成绩是________环.



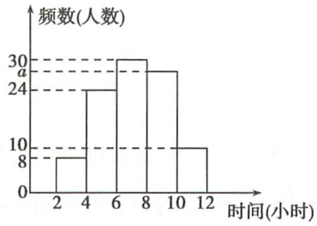
解析 由条形图可知，该队员的平均成绩为$\frac{6\times1+7\times1+8\times2+9\times4+10\times2}{1+1+2+4+2}=8.5$(环).

答案 8.5

知识3 中位数

中位数	将一组数据按大小顺序排列后，处于中间位置的数叫做这组数据的中位数.如果数据的个数为奇数，那么处于中间位置的一个数据是这组数据的中位数；如果数据的个数为偶数，那么处于中间位置的两个数据的平均数是这组数据的中位数
温馨提示	(1)求中位数的步骤：①排序：由小到大或由大到小；②结果：如果数据的个数 n 是奇数，那么第$\frac{n+1}{2}$个数是中位数；如果数据的个数 n 是偶数，那么第$\frac{n}{2}$个和第$\frac{n}{2}+1$ 个数的平均数是中位数. (2)一组数据的中位数是唯一的，它可能是原数据中的一个数，也可能不是原数据中的数. (3)由一组数据的中位数可以知道中位数以上和以下的数据各占一半

例3 某校为了解全校学生“五一”假期参加社团活动的情况，抽查了100名学生，统计他们假期参加社团活动的时间，绘成频数分布直方图(如图)，则参加社团活动时间的中位数所在的范围是　　(　　)



A.4~6小时　　B.6~8小时

C.8~10小时　　D.不能确定

解析　100个数据，排序后，中间的两个数为第50个数和第51个数，而第50个数和第51个数都落在第三组，所以参加社团活动时间的中位数所在的范围为6~8小时.故选B.

答案　B

例4 某校开展了主题为“青春梦想”的艺术作品征集活动，从八年级五个班收集到作品的数量(单位：件)分别为42，50，45，46，50；从九年级六个班收集到作品的数量(单位：件)分别为43，50，47，50，49，40，则八、九年级这两组数据的中位数分别是________.

解析　八年级的这组数据从小到大排列为42，45，46，50，50，

处在最中间的数是46，所以这组数据的中位数是46.

九年级的这组数据从小到大排列为40，43，47，49，50，50.

处在最中间的有两个数据47和49，它们的平均数是48，所以这组数据的中位数是48.

答案　46，48

知识4　众数

众数就是一组数据中出现次数最多的数据.一组数据中，众数可能不止一个，它同平均数、中位数一样，都反映了一组数据的集中趋势.

温馨提示

①如果一组数据中有若干个数据的频数一样，都是最大的，那么这若干个数据都是这组数据的众数，即一组数据的众数可以不唯一.

②一组数据的众数一定出现在这组数据中，众数是一组数据中出现次数最多的数据，而不是数据出现的次数.

③众数的大小只与这组数据中部分数据有关，当一组数据中有个别数据多次重复出现，以至于其他数据的作用显得相对较小时，众数可以在某种意义上代表这组数据的整体情况.

例5 求下列各组数据的众数：

(1)5，6，8，16，16；

(2)75，90，90，85，90，95，85，100，85；

(3)1，2，3，4，5.

解析　(1)这组数据中，16出现了两次，次数最多，则众数是16.

(2)这组数据中，85和90各出现了3次，次数最多，则众数是85和90.

(3)这组数据中，各个数据出现的次数相同，则众数不存在.

知识5　平均数、中位数、众数的优缺点

	优点	缺点	联系
平均数	平均数能充分利用各数据提供的信息，在实际生活中常用样本的平均数估计总体的平均数	在计算平均数时，所有的数据都参与运算，所以它易受极端值的影响	平均数、中位数和众数都是描述一组数据的集中趋势的特征数
中位数	中位数不受个别偏大或偏小数据的影响，当一组数据中的个别数据变动较大时，一般用中位数来描述数据的集中趋势	不能充分地利用各数据的信息	
众数	众数考察的是各数据所出现的频数，其大小只与部分数据有关，当一组数据中某些数据多次重复出现时，众数往往更能反映问题	当各数据重复出现的次数大致相等时，它往往就没有什么特别意义	

方法清单

方法1　平均数、中位数、众数的计算方法

方法2　加权平均数的应用方法

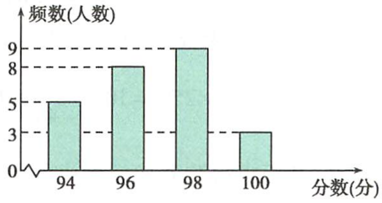
方法3　选择合适的统计量解决问题

方法4　求统计图表中的“三数”的方法

方法1　平均数、中位数、众数的计算方法

对一组数据的平均数、中位数、众数要严格按照其定义进行计算，特别是中位数的计算，要注意数据个数是奇数还是偶数，数据个数为奇数时，其中位数是排序后处于中间位置的数，数据个数为偶数时，其中位数是排序后中间两个数的平均数.一组数据的平均数、中位数只有一个，而众数可能不止一个.

例 1 在光明中学组织的全校师生迎“五四”诗词大赛中，来自不同年级的 25 名参赛同学的得分情况如图所示. 这些成绩的中位数和众数分别是 ( )



A.96 分、98 分　　B.97 分、98 分

C.98 分、96 分　　D.97 分、96 分

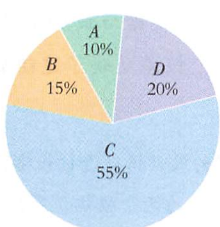
解析 由题图可知，得 98 分的人数最多，有 9 人，故这些成绩的众数是 98 分；由题图可知，得 100 分的有 3 人，得 98 分的有 9 人，得 96 分的有 8 人，得 94 分的有 5 人，把这些成绩按从小到大的顺序排列后，最中间（第 13 个数）的成绩是 96 分，所以中位数是 96 分，故选 A.

答案 A

方法 2 加权平均数的应用方法

在实际问题中，如果一组数据的“重要程度”不相同，求其平均数需采用加权平均数的计算方法. 计算时要根据所给数据的特征，正确识别数据的“重要程度”，进而利用加权平均数进行进一步的分析与决策.

例 2 某超市销售 A,B,C,D 四种矿泉水，它们的单价依次是 5 元、3 元、2 元、1 元. 某天的销售情况如图所示，则这天销售的矿泉水的平均单价是 ( )



A.1.95 元　　B.2.15 元　　C.2.25 元　　D.2.75 元

解析 这天销售的矿泉水的平均单价为 $5\times10\%+3\times15\%+2\times55\%+1\times20\%=2.25$（元），故选 C.

答案 C

例 3 某市某学校招聘数学教师，本次招聘进行专业技能笔试和课堂教学展示两项考核，这两项考核的满分均为 100 分，学校将这两项考核的得分按一定的比例计算出总成绩. 经统计，参加考核的 4 名考生的两项考核的得分如下表所示：

考生序号	1	2	3	4
专业技能笔试	90	70	86	75
课堂教学展示	70	90	80	86

经过计算，1 号考生的总成绩为 78 分.

(1) 求专业技能笔试得分和课堂教学展示得分分别占总成绩的百分比；

(2) 若学校录取总成绩最高的考生，通过计算说明 4 名考生中哪一名考生会被录取.

解析 (1) 设专业技能笔试得分占总成绩的百分比为 a，则课堂教学展示得分占总成绩的百分比为 $1-a$，根据题意，得 $90a+70(1-a)=78$，解得 $a=40\%$，所以 $1-a=60\%$，故专业技能笔试得分占总成绩的 40%，课堂教学展示得分占总成绩的 60%.

(2) 根据 (1) 中的比例计算得，

2 号考生的总成绩为 $70\times40\%+90\times60\%=82$（分），

3 号考生的总成绩为 $86\times40\%+80\times60\%=82.4$（分），

4 号考生的总成绩为 $75\times40\%+86\times60\%=81.6$（分），

$\therefore$ 3 号考生会被录取.

方法 3 选择合适的统计量解决问题

当一组数据中某个数据特别大时，平均数不能反映这组数据的集中趋势，应选用中位数或众数来代表该组数据的“平均水平”. 平均数、中位数和众数都是用来刻画数据平均水平的统计量，平均数常用于表示统计对象的一般水平，中位数表示这组数据的中等水平，而众数刻画了数据中出现次数最多的情况.

例 4 下表是随机抽取的某公司部分员工的月收入资料：

月收入/元	45 000	18 000	10 000	5 500	5 000	3 400	3 000	2 000
人数	1	1	1	3	6	1	11	2

(1) 请计算以上样本的平均数和中位数；

(2) 甲、乙两人分别用样本平均数和中位数来推断公司全体员工月收入水平，请你写出甲、乙两人的推断结论.

解析 (1) 样本的平均数 $\bar{x}=$

$$\frac{45\ 000+18\ 000+10\ 000+5\ 500\times3+5\ 000\times6+3\ 400+3\ 000\times11+2\ 000\times2}{1+1+1+3+6+1+11+2}$$

$=6\ 150$（元）.

这组数据共有 26 个，按月收入从高到低排列后，第 13、14 个数据分别是 3 400、3 000，所以样本的中位数是 $\frac{3\ 400+3\ 000}{2}=3\ 200$（元）.

(2) 甲：由样本平均数为 6 150 元推断公司全体员工月平均收入大约为 6 150 元.

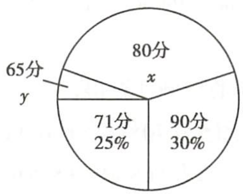
乙：由样本中位数为 3 200 元推断公司全体员工大约有一半的员工月收入超过 3 200 元，有一半的员工月收入不足 3 200 元.

方法 4 求统计图表中的“三数”的方法

表示数据的主要工具是统计图表，条形统计图、扇形统计图、折线统计图是三种常用的统计图. 三种统计图表示数据时都有形象直观、见图知意的优点. 利用各数据提供的信息，便可计算平均数和中位数；根据各数据所出现的频数，就可确定众数.

例 5 某班同学数学测验成绩的统计表和扇形统计图如下：

人数	a	16	b	2
成绩/分	90	80	71	65



则 $a=$______，$b=$______；$x=$______，$y=$______；测验成绩的中位数是______，众数是______，平均数是________.

解析 由题表可知，成绩为 80 分和 65 分的共有 18 人，占总人数的 $1-25\%-30\%=45\%$，∴ 班级总人数为 $18\div45\%=40$，∴ $a=40\times30\%=12$，$b=40\times25\%=10$，$x=\frac{16}{40}\times100\%=40\%$，$y=\frac{2}{40}\times100\%=5\%$，从而求得测验成绩的中位数为 80 分，众数为 80 分，平均数为 80 分.

答案 12；10；40%；5%；80 分；80 分；80 分

29.2 数据的波动

知识清单

知识 方差

知识 方差

1.在一组数据 $x_1,x_2,\cdots,x_n$ 中，各数据与它们的平均数的差的平方的平均数叫做这组数据的方差，通常用 s^2 表示，即

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

2.方差的计算方法有如下三种：

①定义法：就是利用上面方差的定义公式计算.

②原始数据计算法：当数据较小时，可直接利用原始数据进行计算：$s^2=\frac{1}{n}[(x_1^2+x_2^2+\cdots+x_n^2)-n\bar{x}^2]$.

③新数据计算法：当数据较大且比较集中时，可以依照简化平均数的计算方法，将每个数据同时减去与它们的平均数接近的常数 a，得到一组新数据：$x_1'=x_1-a$，$x_2'=x_2-a$，……，$x_n'=x_n-a$，那么 $s^2=\frac{1}{n}[(x_1'^2+x_2'^2+\cdots+x_n'^2)-n\overline{x'}^2]$.

3.求方差的步骤：求方差的步骤可概括为一均，二差，三方，四再均.即第一步求原数据的平均数，第二步求原数据中各数据与平均数的差，第三步求所得各个差的平方，第四步求所得各平方数的平均数.

温馨提示

①方差反映的是数据在它的平均数附近波动的情况，是用来衡量一组数据波动大小的量.

②方差能够反映所有数据的信息，因而在刻画数据波动情况时比极差更准确.方差越大，数据波动越大；方差越小，数据波动越小.只有当两组数据的平均数相等或接近时，才能用方差比较它们波动的情况.

③一组数据的每个数据都变为原来的 k 倍，则所得的一组新数据的方差将变为原数据方差的 k^2 倍.

④方差的单位是原数据单位的平方，在具体使用时可不标注单位.

例 已知一个样本的方差 $s^2=\frac{1}{11}[(x_1-6)^2+(x_2-6)^2+\cdots+(x_{11}-6)^2]$，则这个样本的容量是________，样本的平均数是________.

解析 样本方差的公式为 $s^2=\frac{1}{n}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]$，其中 n 为样本数据的个数，$\bar{x}$为平均数，故这个样本的容量为 11，样本的平均数为 6.

答案 11；6

知识拓展

1.极差

一组数据中的最大数据与最小数据的差叫做极差，即极差＝最大值－最小值.极差反映了这组数据的变化范围.

温馨提示

①在对一组数据的波动情况进行粗略估计时，经常用极差.

②极差是由数据中的两个极端值所决定的，当个别极端值远离其他数据时，极差往往不能反映全体数据的实际波动情况.

③极差小，平均数对这组数据的一般水平的代表性就大；极差大，平均数对这组数据的一般水平的代表性就小.

例 在九年级体育测试中，某班参加仰卧起坐测试的一组女生（每组 8 人）的测试成绩如下（单位：次/分）：46，44，45，42，48，46，47，45，则这组数据的极差为 （　　）

A.2　　B.4

C.6　　D.8

解析 ∵ 46，44，45，42，48，46，47，45 中最大的数是 48，最小的数是 42，

∴ 这组数据的极差为 $48-42=6$，故选 C.

答案 C

2.极差与方差的区别与联系

(1)极差与方差的区别

①极差反映的仅仅是数据的变化范围；方差反映的是数据在它的平均数附近波动的情况.

②计算极差时,只需要计算数据的最大值与最小值的差即可,而方差的计算就要复杂得多.方差是一组数据中各个数据与这组数据平均数的差的平方的平均数.

(2)极差与方差的联系

极差、方差都是用来描述一组数据波动情况的,常用来比较两组数据波动的大小,极差、方差越小,波动越小,进而知这组数据比较稳定;极差、方差越大,波动越大,进而知这组数据不稳定.

3.标准差

方差的算术平方根叫做这组数据的标准差,用“s”表示,即

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

温馨提示

①标准差的数量单位与原数据一致.

②标准差也用来描述一组数据波动情况,常用来比较两组数据波动的大小.

方法清单

方法1 方差的性质的应用方法 ★★☆

方法2 极差、方差的应用方法 ★★★

方法3 利用方差进行决策的方法 ★★☆

方法 1 方差的性质的应用方法

若题目中已知一组数据的方差,而另一组数据又是由原始数据中每个数据通过一定四则运算得到的,则可用方差的性质求解另一组数据的方差,例如:一组数据 $x_1,x_2,\cdots,x_n$ 的方差为 s^2,则(1)$x_1\pm b,x_2\pm b,\cdots\cdots,x_n\pm b$ 的方差为 s^2;(2)$ax_1\pm b,ax_2\pm b,\cdots\cdots,ax_n\pm b$ 的方差为 a^2s^2.

例1 (1)如果将一组数据中的每一个数据都加上同一个非零常数,那么这组数据的 (　　)

A.平均数和方差都不变　　B.平均数不变,方差改变

C.平均数改变,方差不变　　D.平均数和方差都改变

(2)一组数据的方差为 s^2,将这组数据中的每个数都除以2,所得新数据的方差是 (　　)

A.$\frac{1}{2}s^2$　　B.$2s^2$

C.$\frac{1}{4}s^2$　　D.$4s^2$

(3)已知一组数据 x_1,x_2,x_3,x_4,x_5 的平均数是2,方差是 $\frac{1}{3}$,那么另一组数据 $3x_1-2,3x_2-2,3x_3-2,3x_4-2,3x_5-2$ 的平均数和方差分别为 (　　)

A.$2,\frac{1}{3}$　　B.2,1

C.$4,\frac{2}{3}$　　D.4,3

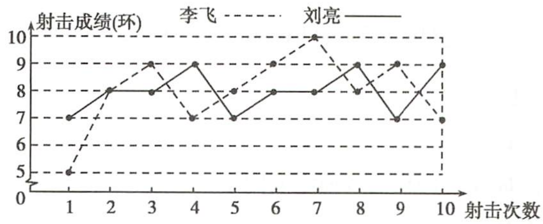
解析 若 $x_1,x_2,\cdots,x_n$ 的平均数为$\bar{x}$,方差为 s^2,则(1)$x_1\pm a,x_2\pm a,\cdots\cdots,x_n\pm a$ 的平均数为$\bar{x}\pm a$,方差不变;(2)$kx_1,kx_2,\cdots,kx_n$ 的平均数为 $k\bar{x}$,方差为 k^2s^2;(3)$kx_1+a,kx_2+a,\cdots\cdots,kx_n+a$ 的平均数为 $k\bar{x}+a$,方差为 k^2s^2.

答案 (1)**C**　(2)**C**　(3)**D**

方法 2 极差、方差的应用方法 ★★★

极差反映数据的波动范围,计算方便.方差反映数据的稳定性,方差越大,说明其稳定性越差;方差越小,说明其稳定性越好.

例2 根据李飞与刘亮射击训练的成绩绘制了如图所示的折线统计图.



根据折线统计图所提供的信息,若要推荐一位成绩较稳定的选手去参赛,应推荐 (　　)

A.李飞或刘亮

B.李飞

C.刘亮

D.无法确定

解析 李飞的成绩(单位:环)为5、8、9、7、8、9、10、8、9、7,

则李飞成绩的平均数为$\frac{5+7\times2+8\times3+9\times3+10}{10}=8$(环),

$\therefore$ 李飞成绩的方差为$\frac{1}{10}\times[(5-8)^2+2\times(7-8)^2+3\times(8-8)^2+3\times(9-8)^2+(10-8)^2]=1.8$;

刘亮的成绩(单位:环)为7、8、8、9、7、8、8、9、7、9,

则刘亮成绩的平均数为$\frac{7\times3+8\times4+9\times3}{10}=8$(环),

$\therefore$ 刘亮成绩的方差为$\frac{1}{10}\times[3\times(7-8)^2+4\times(8-8)^2+3\times(9-8)^2]=0.6$.

$\because 0.6<1.8$,

$\therefore$ 应推荐刘亮,故选C.

答案 C

例 3 下表是甲、乙两名同学近五次数学测试（满分均为 100 分）的成绩统计表.

	第一次	第二次	第三次	第四次	第五次
甲	90	88	92	94	91
乙	90	91	93	94	92

根据上表中数据，成绩较好且比较稳定的同学是________.

思路分析 一组数据的平均数反映这组数据的平均水平，平均分越高成绩越好.方差反映一组数据在其平均数附近波动的大小，方差越大，数据波动就越大，越不稳定；方差越小，数据波动就越小，越稳定.

解析 $\bar{x}_{甲}=\dfrac{90+88+92+94+91}{5}=91$（分），

$\bar{x}_{乙}=\dfrac{90+91+93+94+92}{5}=92$（分），

$s_{甲}^2=\dfrac{1}{5}\times[(90-91)^2+(88-91)^2+(92-91)^2+(94-91)^2+(91-91)^2]=4$，

$s_{乙}^2=\dfrac{1}{5}\times[(90-92)^2+(91-92)^2+(93-92)^2+(94-92)^2+(92-92)^2]=2$.

$\because \bar{x}_{甲}<\bar{x}_{乙}, s_{乙}^2<s_{甲}^2$，

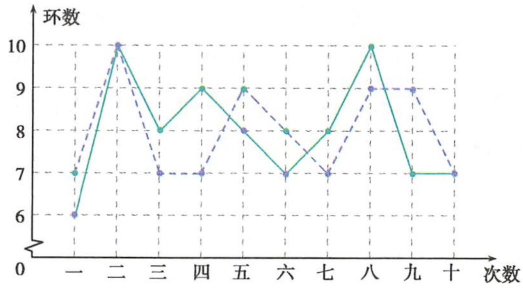
$\therefore$ 成绩较好且比较稳定的同学是乙.

答案 乙

方法 3 利用方差进行决策的方法

在分析数据时，往往要求数据的平均数，当数据的平均水平一致时，为了更好地根据统计结果进行合理的判断和预测，我们往往会根据方差来判断数据的稳定性，从而得到正确的决策.

例 4 甲、乙两名射击运动员进行射击比赛，两人在相同条件下各射击 10 次，射击的成绩如图所示.



（实线表示甲，虚线表示乙）

根据图中信息，回答下列问题：

（1）求甲的平均数与乙的中位数；

（2）分别计算甲、乙成绩的方差，并从计算结果分析哪位运动员的射击成绩更稳定.

解析 （1）甲的平均数为$\dfrac{6+10+8+9+8+7+8+10+7+7}{10}=8$，乙的中位数是 7.5.

（2）乙的平均数为$\dfrac{1}{10}\times(7+10+7+7+9+8+7+9+9+7)=8$，

$s_{甲}^2=\dfrac{1}{10}\times[(6-8)^2+2\times(10-8)^2+3\times(8-8)^2+(9-8)^2+3\times(7-8)^2]=1.6$，

$s_{乙}^2=\dfrac{1}{10}\times[5\times(7-8)^2+(8-8)^2+3\times(9-8)^2+(10-8)^2]=1.2$.

$\because s_{乙}^2<s_{甲}^2$，

$\therefore$ 乙运动员的射击成绩更稳定.