高一年级数学第57课时总体离散程度估计学习指南

一、学习目标

(1)结合实例，能用样本估计总体的离散程度参数（标准差、方差、极差）

(2)理解离散程度参数的统计含义.

二、学法指导

平均数、中位数和众数为我们提供了一组数据的集中趋势的信息，这是概括一组数据的特征的有效方法.但仅知道集中趋势的信息，很多时候还不能使我们做出有效决策，例如下面的问题.

任务一：提出刻画数据离散程度的需要

问题1有两位射击运动员在一次射击测试中各射靶10次,每次命中的环数如下：

甲	7	8	7	9	5	4	9	10	7	4
乙	9	5	7	8	7	6	8	6	7	7

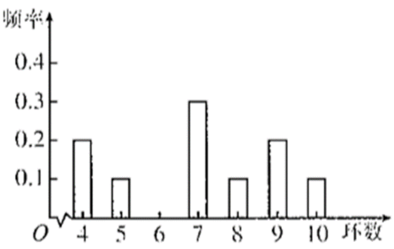
如果你是教练，你如何对两位运动员的射击情况作出评价？

如果这是一次选拔性考核，你应当如何作出选择？

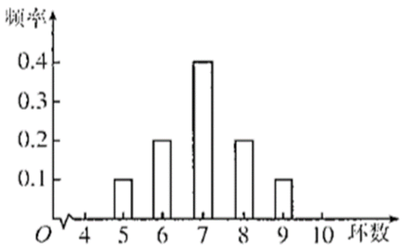
通过计算可得

甲：　平均数7　　中位数7　　众数7

乙：　平均数7　　中位数7　　众数7



(甲)



(乙)

从甲、乙成绩的频率分布直方图看：甲的成绩比较分散，乙的成绩相对集中。即甲的成绩波动幅度较大，乙的成绩比较稳定。

如何度量成绩的这种差异呢？

一种简单的度量数据离散程度的方法就是用极差.根据甲、乙运动员的10次射击成绩，可以得到

甲命中环数的极差=10-4=6

乙命中环数的极差=9-5=4

可以发现甲的成绩波动范围比乙的大.极差在一定范围内刻画了数据的离散

程度.

但因为极差只使用了数据中最大、最小两个值的信息，对其他数据的取值情况没有涉及，所以极差所含的信息量很少.

我们知道，如果射击成绩很稳定，那么大多数的射击成绩离平均成绩不会太远；相反，如果射击的成绩波动幅度很大，那么大多数的射击成绩离平均成绩会比较远. 因此，我们可以通过这两组射击成绩与它们的平均成绩的“平均距离”来度量成绩的波动幅度.

思考：如何定义“平均距离”？

任务二：引入方差的概念和计算方法

假设一组数据是 $x_1, x_2, x_3, \cdots x_n$. 用 $\overline{x}$ 表示这组数据的平均数.我们用每个数据与平均数的差的绝对值作为“距离”，即

$$\left|x_i - \overline{x}\right| \qquad (i = 1,2,3,\cdots n)$$

作为 x_i 到 $\overline{x}$ 的距离.可以得到这组数据到 $\overline{x}$ 的“平均距离”为

$$\frac{1}{n}\sum_{i=1}^{n}\left|x_i - \overline{x}\right|.$$

为了避免式中含绝对值，通常改用平方来代替，即

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2. \text{（1）}$$

我们称（1）式为这组数据的方差.

有时为了计算方差的方便，我们还把方差写成以下形式 $\frac{1}{n}\sum_{i=1}^{n}x_i^{\ 2} - \overline{x}^2$.

由于方差的单位是原始数据单位的平方，与原始数据不一致.为了使二者单位一致，我们对方差开平方，取它的算术平方根，即

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}. \text{（2）}$$

我们称（2）式为这组数据的标准差.

如果总体中所有个体的变量值分别为 $Y_1, Y_2, Y_3, \cdots Y_N$, 总体平均数为 $\overline{Y}$, 则称

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2$$ 为总体方差， $S = \sqrt{S^2}$ 为总体标准差.

与总体均值类似，总体方差也可以写成加权形式.

如果总体的 N 个变量值中，不同的值共有 $k(k\le N)$ 个，不妨记为 $Y_1,Y_2,Y_3,\cdots Y_k$，其中 Y_i 出现的频数为 $f_i(i=1,2,3,\cdots k)$，则总体方差为

$$S^2=\frac{1}{N}\sum_{i=1}^{k}f_i\left(Y_i-\overline{Y}\right)^2.$$

如果一个样本中个体的变量值分别为 $y_1,y_2,y_3,\cdots y_n$，样本平均数为 $\overline{y}$，则称

$s^2=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\overline{y}\right)^2$ 为**样本方差**，$s=\sqrt{s^2}$ 为**样本标准差**.

任务三:方差的应用——总体离散程度的估计

例 在对树人中学高一年级学生身高的调查中，采用样本量比例分配的分层随机抽样，如果不知道样本数据，只知道男生抽取了 23 人，其平均数和方差分别为 170.6 和 12.59，抽取了女生 27 人，其平均数和方差分别 160.6 和 38.62，你能由这些数据计算出总样本的方差，并对高一年级全体学生的方差作出估计吗？

解： 把男生样本记为 $x_1,x_2,x_3,\cdots x_{23}$，其平均数记为 $\overline{x}$，方差记为 $s_x{}^2$；

把女生样本记为 $y_1,y_2,y_3,\cdots y_{27}$，其平均数记为 $\overline{y}$，方差记为 $s_y{}^2$；

把总样本数据的平均数记为 $\overline{z}$，方差记为 s^2.

根据方差的定义，总样本的方差为

$$s^2=\frac{1}{50}\left[\sum_{i=1}^{23}\left(x_i-\overline{z}\right)^2+\sum_{j=1}^{27}\left(y_i-\overline{z}\right)^2\right]$$

$$=\frac{1}{50}\left[\sum_{i=1}^{23}\left(x_i-\overline{x}+\overline{x}-\overline{z}\right)^2+\sum_{j=1}^{27}\left(y_i-\overline{y}+\overline{y}-\overline{z}\right)^2\right]$$

由 $\sum_{i=1}^{23}\left(x_i-\overline{x}\right)=\sum_{i=1}^{23}x_i-23\overline{x}=0$，可得 $\sum_{i=1}^{23}2\left(x_i-\overline{x}\right)\left(\overline{x}-\overline{z}\right)=2\left(\overline{x}-\overline{z}\right)\sum_{i=1}^{23}\left(x_i-\overline{x}\right)=0$

同理可得 $\sum_{i=1}^{27}2\left(y_i-\overline{y}\right)\left(\overline{y}-\overline{z}\right)=0$

因此 $s^2=\frac{1}{50}\left[\sum_{i=1}^{23}\left(x_i-\overline{x}\right)^2+\sum_{i=1}^{23}\left(\overline{x}-\overline{z}\right)^2+\sum_{j=1}^{27}\left(y_i-\overline{y}\right)^2+\sum_{j=1}^{27}\left(\overline{y}-\overline{z}\right)^2\right]$

$=\frac{1}{50}\left\{23\left[s_x{}^2+\left(\overline{x}-\overline{z}\right)^2\right]+27\left[s_y{}^2+\left(\overline{y}-\overline{z}\right)^2\right]\right\}$ ①

由 $\overline{x}=170.6,\overline{y}=160.6$，根据按比例分配分层随机抽样总样本平均数与各层样本平均数的关系，可得总样本平均数为

$$\bar{z}=\frac{23}{23+27}\bar{x}+\frac{27}{23+27}\bar{y}=\frac{23\times170.6+27\times160.0}{50}=165.2$$

把已知的男生、女生样本平均数和方差的取值代入①，可得 $s^2=51.4862$

我们可以计算出总样本的方差为 51.4862，并据此估计高一年级学生身高的总体方差为 51.4862.