## 14.4.2　用样本估计总体的离散程度参数

学习目标　1.理解样本数据方差、标准差的意义，会计算方差、标准差.2.会用样本的基本数字特征(平均数、标准差)估计总体的基本数字特征.3.体会用样本估计总体的思想．

知识点一　极差

1．定义：一组数据的最大值与最小值的差．

2．作用：极差较大，数据点较分散；极差较小，数据点较集中．

知识点二　方差、标准差

1．方差：设一组样本数据*x*1，*x*2，…，*xn*，其平均数为，则称*s*2＝(*xi*－)2为这个样本的方差，简称样本方差．

2．标准差：方差的算术平方根*s*＝为样本的标准差，简称样本标准差．

3．标准差(或方差)越小，数据越稳定在平均数附近．*s*＝0时，每一组样本数据均为.

1．数据2,3,4,5的标准差是数据4,6,8,10的标准差的一半．(　√　)

2．方差与标准差具有相同的单位．(　×　)

3．方差的值越小，数据的离散程度越小．(　√　)

4．如果一组数中每个数减去同一个非零常数，则这组数的平均数改变，方差不变(　√　)

一、方差、标准差的计算

例1　从甲、乙两种玉米中各抽10株，分别测得它们的株高：

甲：25,41,40,37,22,14,19,39,21,42；

乙：27,16,44,27,44,16,40,40,16,40.

试计算甲、乙两组数据的方差和标准差．

解　甲＝×(25＋41＋40＋37＋22＋14＋19＋39＋21＋42)＝30，

*s*＝×[(25－30)2＋(41－30)2＋…＋(42－30)2]＝104.2，

*s*甲＝≈10.208.

乙＝×(27＋16＋44＋27＋44＋16＋40＋40＋16＋40)＝31，

同理*s*＝128.8，*s*乙＝≈11.349.

反思感悟　方差的计算方法

(1)*s*2＝[(*x*＋*x*＋…＋*x*)－*n*·2]；

*s*2＝(*x*＋*x*＋…＋*x*)－2.

(2)用定义的公式计算方差的一般步骤

①先求出样本平均数；

②再计算一组差：*xi*－(*i*＝1,2，…，*n*)；

③计算②中差的平方，得到一组新的数据：(*x*1－)2，(*x*2－)2，…，(*xn*－)2；

④计算③中这组新数据的平均数，即为所求的方差*s*2，即*s*2＝(*xi*－)2.

跟踪训练1　已知一个样本为1,3,2,5，*x*，它的平均数是3，则这个样本的标准差是多少？

解　方法一　∵＝＝3，∴*x*＝4.

由方差公式得，*s*2＝[(1－3)2＋(3－3)2＋(2－3)2＋(5－3)2＋(4－3)2]＝2，∴*s*＝.

方法二　∵＝＝3，∴*x*＝4，

由方差公式的变形公式得，*s*2＝(12＋32＋22＋52＋42)－32＝2，∴*s*＝.

二、方差的性质

例2　设数据*x*1，*x*2，…，*xn*的方差为*s*2，求下列各组数据的方差．

(1)*x*1＋*b*，*x*2＋*b*，…，*xn*＋*b*；

(2)*ax*1，*ax*2，…，*axn*；

(3)*ax*1＋*b*，*ax*2＋*b*，…，*axn*＋*b*.

解　设数据*x*1，*x*2，…，*xn*的平均数为，

则数据*x*1＋*b*，*x*2＋*b*，…，*xn*＋*b*的平均数为＋*b*，

数据*ax*1，*ax*2，…，*axn*的平均数为*a*，

数据*ax*1＋*b*，*ax*2＋*b*，…，*axn*＋*b*的平均数为*a*＋*b*，

设数据*x*1＋*b*，*x*2＋*b*，…，*xn*＋*b*的方差为*s*，

数据*ax*1，*ax*2，…，*axn*的方差为*s*，

数据*ax*1＋*b*，*ax*2＋*b*，…，*axn*＋*b*的方差为*s*.

(1)*s*＝[(*x*1＋*b*－－*b*)2＋(*x*2＋*b*－－*b*)2＋…＋(*xn*＋*b*－－*b*)2]

＝[(*x*1－)2＋(*x*2－)2＋…＋(*xn*－)2]＝*s*2.

(2)*s*＝[(*ax*1－*a*)2＋(*ax*2－*a*)2＋…＋(*axn*－*a*)2]

＝*a*2·[(*x*1－)2＋(*x*2－)2＋…＋(*xn*－)2]＝*a*2*s*2.

(3)*s*＝[(*ax*1＋*b*－*a*－*b*)2＋(*ax*2＋*b*－*a*－*b*)2＋…＋(*axn*＋*b*－*a*－*b*)2]

＝[(*ax*1－*a*)2＋(*ax*2－*a*)2＋…＋(*axn*－*a*)2]

＝*a*2*s*2.

反思感悟　方差的性质

(1)数据*x*1，*x*2，…，*xn*与数据*x*1＋*b*，*x*2＋*b*，…，*xn*＋*b*的方差相等．

(2)若*x*1，*x*2，…，*xn*的方差为*s*2，则*ax*1，*ax*2，…，*axn*的方差为*a*2*s*2.

(3)若*x*1，*x*2，…，*xn*的方差为*s*2，则*ax*1＋*b*，*ax*2＋*b*，…，*axn*＋*b*的方差为*a*2*s*2.

利用这些性质可比较方便地求一些数据的方差．

跟踪训练2　(1)已知一组数据*x*1，*x*2，…，*x*8的平均数是2，方差为6，则数据*x*1－1，*x*2－1，…，*xn*－1的平均数是\_\_\_\_\_\_\_\_，方差是\_\_\_\_\_\_\_\_．

(2)已知一组数据*x*1，*x*2，…，*xn*的平均数是－2，方差是4，则数据2*x*1＋3,2*x*2＋3，…，2*xn*＋3的平均数是\_\_\_\_\_\_\_\_，方差是\_\_\_\_\_\_\_\_．

答案　(1)1　6　(2)－1　16

三、方差、标准差的应用

例3　甲、乙两名战士在相同条件下各打靶10次，每次命中的环数分别为：

甲：8,6,7,8,6,5,9,10,4,7；

乙：6,7,7,8,6,7,8,7,9,5.

(1)分别计算以上两组数据的平均数；

(2)分别求出两组数据的方差和标准差；

(3)根据计算结果，估计两名战士的射击情况．若要从这两人中选一人参加射击比赛，选谁去合适？

(4)估计两名战士射击环数落在区间(－*s*，＋*s*)内的百分比是多少．

解　(1)甲＝×(8＋6＋7＋8＋6＋5＋9＋10＋4＋7)＝7，

乙＝×(6＋7＋7＋8＋6＋7＋8＋7＋9＋5)＝7.

(2)由方差公式*s*2＝[(*x*1－)2＋(*x*2－)2＋…＋(*xn*－)2]，得*s*＝3，*s*＝1.2.

故*s*甲≈1.7，*s*乙≈1.1.

(3)甲＝乙，说明甲、乙两战士的平均水平相当．

又*s*＞*s*，说明甲战士射击情况波动大．

因此，乙战士比甲战士射击情况稳定．从成绩的稳定性考虑，应选择乙参加比赛．

(4)对于甲，样本数据落在(－*s*，＋*s*)，即(5.3,8.7)内的有6个，占60%.对于乙，样本数据落在(－*s*，＋*s*)，即(5.9,8.1)内的有8个，占80%.

反思感悟　在实际问题中，仅靠平均数不能完全反映问题，还要研究方差，方差描述了数据相对平均数的离散程度，在平均数相同的情况下，方差越大，离散程度越大，数据波动性越大，稳定性越差；方差越小，数据越集中，越稳定．

跟踪训练3　某化肥厂有甲、乙两个车间包装肥料，在自动包装传送带上每隔30分钟抽取一包产品，称其质量，分别记录抽查数据如下(单位：kg)：

甲：102　101　99　98　103　98　99

乙：110　115　90　85　75　115　110

试计算甲、乙两个车间产品质量的平均数与方差，并说明哪个车间产品比较稳定．

解　甲＝(102＋101＋99＋98＋103＋98＋99)＝100；

乙＝(110＋115＋90＋85＋75＋115＋110)＝100；

*s*＝[(102－100)2＋(101－100)2＋(99－100)2＋(98－100)2＋(103－100)2＋(98－100)2＋(99－100)2]＝(4＋1＋1＋4＋9＋4＋1)≈3.43；

*s*＝[(110－100)2＋(115－100)2＋(90－100)2＋(85－100)2＋(75－100)2＋(115－100)2＋(110－100)2]＝(100＋225＋100＋225＋625＋225＋100)≈228.57.

所以*s*＜*s*，故甲车间产品较稳定．

1．下列说法正确的是(　　)

A．在两组数据中，平均数较大的一组方差较大

B．平均数反映数据的集中趋势，方差则反映数据离平均数的波动大小

C．方差的求法是求出各个数据与平均数的差的平方后再求和

D．在记录两个人射击环数的两组数据中，方差大的表示射击水平高

答案　B

解析　A中平均数和方差是数据的两个特征，不存在这种关系；C中求和后还需取平均数；D中方差越大，射击越不平稳．

2．一组数据如下(不完全按大小排列)：2,4,4,5,5,6,7,8,9,11，*x*.已知这组数据的平均数为6，则这组数据的方差为(　　)

A．6 B.

C．66 D．6.5

答案　A

解析　∵这组数据的平均数为×(2＋4＋4＋5＋5＋6＋7＋8＋9＋11＋*x*)＝×(61＋*x*)＝6，∴*x*＝5，则该组数据的方差为*s*2＝＝6.

3．甲、乙、丙、丁四名射击手在选拔赛中所得的平均环数及其方差*s*2如下表所示，则选择决赛的最佳人选应是(　　)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 甲 | 乙 | 丙 | 丁 |
|  | 7 | 8 | 8 | 7 |
| *s*2 | 6.3 | 6.3 | 7 | 8.7 |

A.甲 B．乙 C．丙 D．丁

答案　B

解析　∵乙＝丙>甲＝丁，且*s*＝*s*<*s*<*s*，故应选择乙进入决赛．

4．已知某7个数的平均数为4，方差为2，现加入一个新数据4，此时这8个数的平均数为，方差为*s*2，则(　　)

A.＝4，*s*2＝2 B.＝4，*s*2>2

C.＝4，*s*2<2 D.>4，*s*2<2

答案　C

解析　根据题意有＝＝4，

而*s*2＝<2.

5．样本中共有五个个体，其值分别为*a*,0,1,2,3，若该样本的平均数为1，则样本方差为\_\_\_\_\_\_\_\_．

答案　2

解析　由题意知(*a*＋0＋1＋2＋3)＝1，解得*a*＝－1，所以样本方差为*s*2＝[(－1－1)2＋(0－1)2＋(1－1)2＋(2－1)2＋(3－1)2]＝2.

1．知识清单：标准差、方差的计算及应用．

2．方法归纳：数据分析统计．

3．常见误区：混淆方差(标准差)的意义，导致出错．

1．已知一个样本中的数据为1,2,3,4,5，则该样本的标准差为(　　)

A. B. C．2 D.

答案　A

解析　∵样本容量*n*＝5，

∴＝(1＋2＋3＋4＋5)＝3，

∴*s*＝＝.

2．若1,2,3，*x*的平均数是5，而1,3,3，*x*，*y*的平均数是6，则1,2,3，*x*，*y*的方差是(　　)

A．24.52 B．24.54

C．24.56 D．24.58

答案　C

解析　由＝5得*x*＝14，

由＝6得*y*＝9.

所以*s*2＝24.56.

3．某高三学生在连续五次月考中的数学成绩(单位：分)为：90,90,93,94,93，则该学生在这五次月考中数学成绩的平均数和方差分别为(　　)

A．92,2.8 B．92,2

C．93,2 D．93,2.8

答案　A

解析　该学生在这五次月考中数学成绩的平均数为＝×(90＋90＋93＋94＋93)＝92，

方差为*s*2＝×[(90－92)2＋(90－92)2＋(93－92)2＋(94－92)2＋(93－92)2]＝2.8.

4．如图，样本*A*和*B*分别取自两个不同的总体，它们的样本平均数分别为*A*和*B*，样本标准差分别为*sA*和*s**B*，则(　　)

A.*A*＞*B*，*sA*＞*sB* B.*A*＜*B*，*sA*＞*sB*

C.*A*＞*B*，*sA*＜*sB* D.*A*＜*B*，*sA*＜*sB*

答案　B

解析　由题图知，*A*组的6个数分别为2.5,10,5,7.5，2.5，10；*B*组的6个数分别为15,10,12.5,10,12.5,10，

显然*A*<*B*.

又由图形可知，*B*组数据的分布比*A*组的均匀，变化幅度不大，故*B*组数据比较稳定，方差较小，从而标准差较小，所以*sA*>*sB*.

5．样本*a*,3,5,7的平均数是*b*，且*a*，*b*是方程*x*2－5*x*＋4＝0的两根，则这个样本的方差是(　　)

A．3 B．4 C．5 D．6

答案　C

解析　*x*2－5*x*＋4＝0的两根是1,4.

当*a*＝1时，*a*,3,5,7的平均数是4；

当*a*＝4时，*a*,3,5,7的平均数不是1.

∴*a*＝1，*b*＝4，则方差*s*2＝×[(1－4)2＋(3－4)2＋(5－4)2＋(7－4)2]＝5.

6．某校甲、乙两个班级各有5名编号为1,2,3,4,5的学生进行投篮练习，每人投10次，投中的次数如下表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 学生 | 1号 | 2号 | 3号 | 4号 | 5号 |
| 甲班 | 6 | 7 | 7 | 8 | 7 |
| 乙班 | 6 | 7 | 6 | 7 | 9 |

则以上两组数据的方差中较小的一组数据的*s*2＝\_\_\_\_\_\_.

答案

解析　由题意知甲＝(6＋7＋7＋8＋7)＝7，

乙＝(6＋7＋6＋7＋9)＝7，

*s*＝[(6－7)2＋…＋(7－7)2]＝，

*s*＝[(6－7)2＋…＋(9－7)2]＝.

∵＜，∴较小的一个*s*2＝.

7．若样本数据*x*1，*x*2，…，*x*10的标准差为8，则数据2*x*1－1,2*x*2－1，…，2*x*10－1的标准差为\_\_\_\_\_\_\_\_．

答案　16

解析　设样本数据*x*1，*x*2，…，*x*10的标准差为*s*，则*s*＝8，

可知数据2*x*1－1,2*x*2－1，…，2*x*10－1的标准差为2*s*＝16.

8．样本容量为10的一组数据，它们的平均数是5，频率条形图如图，则其标准差为\_\_\_\_\_\_\_\_．

答案

解析　由条形图知2与8的个数相等，且多于5的个数，于是这10个数分别为2,2,2,2,5,5,8,8,8,8.

∵＝5，

∴*s*2＝[(2－5)2＋(2－5)2＋(2－5)2＋(2－5)2＋(5－5)2＋(5－5)2＋(8－5)2＋(8－5)2＋(8－5)2＋(8－5)2]＝×8×9＝，∴*s*＝.

9．某校医务室随机抽查了高一10位男同学的体重(单位：kg)如下：

74,71,72,68,76,73,67,70,65,74.

(1)估计高一所有男同学体重数据的平均数、中位数、方差、标准差；

(2)高一10位男同学的体重数据中，位于[－*s*，＋*s*]内的有几个？所占的百分比是多少？

解　(1)这10位男同学的体重数据的平均数＝×(74＋71＋72＋68＋76＋73＋67＋70＋65＋74)＝71.

将这10位男同学的体重数据按从小到大重新排列，得65,67,68,70,71,72,73,74,74,76，位于中间的两个数是71,72，

所以这10位男同学的体重数据的中位数为＝71.5，这10位男同学的体重数据的方差*s*2＝×[(74－71)2＋(71－71)2＋(72－71)2＋(68－71)2＋(76－71)2＋(73－71)2＋(67－71)2＋(70－71)2＋(65－71)2＋(74－712)]＝11，

标准差*s*＝＝.

(2)因为[－*s*，＋*s*]＝[71－，71＋]，

所以数据74,71,72,68,76,73,67,70,65,74中，有7个数据位于区间[71－，71＋]内，所占的百分比为70%.

10．某学校统计教师职称及年龄，中级职称教师的人数为50，其平均年龄为38岁，方差是2，高级职称的教师中有3人58岁，5人40岁，2人38岁，求该校中级职称和高级职称教师年龄的平均数和方差．

解　由已知条件可知高级职称教师的平均年龄为高＝＝45(岁)，

年龄的方差为*s*＝×[3×(58－45)2＋5×(40－45)2＋2×(38－45)2]＝73，

所以该校中级职称和高级职称教师的平均年龄为＝×38＋×45≈39.2(岁)，

该校中级职称和高级职称教师的年龄的方差是

*s*2＝×[2＋(38－39.2)2]＋×[73＋(45－39.2)2]＝20.64.

11．甲、乙两人在一次射击比赛中各射靶5次，两人成绩的频数直方图如图所示，则(　　)

A．甲的成绩的平均数小于乙的成绩的平均数

B．甲的成绩的中位数等于乙的成绩的中位数

C．甲的成绩的方差小于乙的成绩的方差

D．甲的成绩的极差小于乙的成绩的极差

答案　C

解析　由题意可知，甲的成绩为4,5,6,7,8，乙的成绩为5,5,5,6,9，所以甲、乙成绩的平均数均为6，A错；

甲、乙的成绩的中位数分别为6,5，B错；

甲、乙的成绩的方差分别为×[(4－6)2＋(5－6)2＋(6－6)2＋(7－6)2＋(8－6)2]＝2，×[(5－6)2＋(5－6)2＋(5－6)2＋(6－6)2＋(9－6)2]＝，C对；

甲、乙的成绩的极差均为4，D错．

12．某公司10位员工的月工资(单位：元)为*x*1，*x*2，…，*x*10，其平均数和方差分别为和*s*2，若从下月起每位员工的月工资增加100元，则这10位员工下月工资的平均数和方差分别为(　　)

A.，*s*2＋1002 B.＋100，*s*2＋1002

C.，*s*2 D.＋100，*s*2

答案　D

解析　方法一　因为每个数据都加上100，所以平均数也增加100，而离散程度应保持不变，即方差不变．

方法二　由题意知*x*1＋*x*2＋…＋*x*10＝10 ，

*s*2＝[(*x*1－)2＋(*x*2－)2＋…＋(*x*10－)2]，

则所求平均数＝[(*x*1＋100)＋(*x*2＋100)＋…＋(*x*10＋100)]＝(10＋10×100)＝＋100，

所求方差为[(*x*1＋100－)2＋(*x*2＋100－)2＋…＋(*x*10＋100－)2]＝[(*x*1－)2＋(*x*2－)2＋…＋(*x*10－)2]＝*s*2.

13．某市有15个旅游景点，经计算，黄金周期间各个景点的旅游人数平均为20万，标准差为*s*，后来经核实，发现甲、乙两处景点统计的人数有误，甲景点实际为20万，被误统计为15万，乙景点实际为18万，被误统计成23万；更正后重新计算，得到标准差为*s*1，则*s*与*s*1的大小关系为(　　)

A．*s*＝*s*1 B．*s*<*s*1

C．*s*>*s*1 D．不能确定

答案　C

解析　由已知，两次统计所得的旅游人数总数没有变，即两次统计的各景点旅游人数的平均数是相同的，设为，则*s*＝，

*s*1＝，

若比较*s*与*s*1的大小，只需比较(15－)2＋(23－)2与(20－)2＋(18－)2的大小即可，而(15－)2＋(23－)2＝754－76＋22，(20－)2＋(18－)2＝724－76＋22，所以(15－)2＋(23－)2>(20－)2＋(18－)2，从而*s*>*s*1.

14．为了考察某校各班参加课外书法小组的人数，从全校随机抽取5个班级，把每个班级参加该小组的人数作为样本数据，已知样本平均数为7，样本方差为4，且样本数据互不相同，则样本数据中的最大值为\_\_\_\_\_\_\_\_．

答案　10

解析　设5个班级中参加的人数分别为*x*1，*x*2，*x*3，*x*4，*x*5，

则由题意知＝7，

(*x*1－7)2＋(*x*2－7)2＋(*x*3－7)2＋(*x*4－7)2＋(*x*5－7)2＝20，

五个整数的平方和为20，则必为0＋1＋1＋9＋9＝20，

由|*x*－7|＝3可得*x*＝10或*x*＝4.

由|*x*－7|＝1可得*x*＝8或*x*＝6.

由上可知参加的人数分别为4,6,7,8,10，

故最大值为10.

15．由正整数组成的一组数据*x*1，*x*2，*x*3，*x*4，其平均数和中位数都是2，且标准差等于1，则这组数据为\_\_\_\_\_\_\_\_\_\_．(从小到大排列)

答案　1,1,3,3

解析　假设这组数据按从小到大的顺序排列为*x*1，*x*2，*x*3，*x*4，则

∴

又*s*＝

＝

＝＝1，

∴(*x*1－2)2＋(*x*2－2)2＝2.

同理可求得(*x*3－2)2＋(*x*4－2)2＝2.

由*x*1，*x*2，*x*3，*x*4均为正整数，且(*x*1，*x*2)，(*x*3，*x*4)均为方程(*x*－2)2＋(*y*－2)2＝2的解，分析知*x*1，*x*2，*x*3，*x*4应为1,1,3,3.

16．把某校三年级一班40人随机平均分成两组，两组学生一次考试的成绩情况如下表：

|  |  |  |
| --- | --- | --- |
| 　　　统计量组别 | 平均成绩 | 标准差 |
| 第一组 | 90 | 6 |
| 第二组 | 80 | 4 |

求全班学生的平均成绩和标准差．

解　设第一组20名学生的成绩为*xi*(*i*＝1,2，…，20)，

第二组20名学生的成绩为*yi*(*i*＝1,2，…，20)，

依题意有＝(*x*1＋*x*2＋…＋*x*20)＝90，＝(*y*1＋*y*2＋…＋*y*20)＝80，

故全班平均成绩为(*x*1＋*x*2＋…＋*x*20＋*y*1＋*y*2＋…＋*y*20)＝(90×20＋80×20)＝85；

又设第一组学生成绩的标准差为*s*1，第二组学生成绩的标准差为*s*2，则

*s*＝(*x*＋*x*＋…＋*x*－202)，

*s*＝(*y*＋*y*＋…＋*y*－202)(此处，＝90，＝80)，

又设全班40名学生的标准差为*s*，平均成绩为(＝85)，故有*s*2＝(*x*＋*x*＋…＋*x*＋*y*＋*y*＋…＋*y*－402)

＝(20*s*＋202＋20*s*＋202－402)＝(62＋42＋902＋802－2×852)＝51.

即*s*＝.

所以全班学生的平均成绩为85分，标准差为.