

知识点106 用样本估计总体 (P3-23)



01

知识点106 用样本
估计总体

教材知识萃取

样本的数字特征	众数、中位数、平均数、方差、标准差、极差、百分位数
百分位数	一般地,一组数据的第 p 百分位数是这样—个值,它使得这组数据中至少有 $p\%$ 的数据小于或等于这个值,且至少有 $(100-p)\%$ 的数据大于或等于这个值.
分层随机抽样的 样本的平均数与方差 (以层数是2层为例)	若第1层和第2层包含的个体数分别为 M 和 N ,抽取的样本量分别为 m 和 n ,第1层和第2层的样本平均数分别为 \bar{x}, \bar{y} ,方差分别为 s_x^2, s_y^2 ,样本平均数为 \bar{w} ,方差为 s^2 ,则 $\bar{w} = \frac{M}{M+N}\bar{x} + \frac{N}{M+N}\bar{y} = \frac{m}{m+n}\bar{x} + \frac{n}{m+n}\bar{y}$, $s^2 = \frac{m}{m+n}[s_x^2 + (\bar{x} - \bar{w})^2] + \frac{n}{m+n}[s_y^2 + (\bar{y} - \bar{w})^2]$.

教材知识萃取

常用结论

1. 平均数的性质

(1) 若给定一组数据 x_1, x_2, \dots, x_n 的平均数为 \bar{x} , 则 ax_1, ax_2, \dots, ax_n 的平均数为 $a\bar{x}$, $ax_1 + b, ax_2 + b, \dots, ax_n + b$ 的平均数为 $a\bar{x} + b$.

(2) 若两组数据 x_1, x_2, \dots, x_n 和 y_1, y_2, \dots, y_n 的平均数分别是 \bar{x} 和 \bar{y} , 则 $x_1 + y_1, x_2 + y_2, \dots, x_n + y_n$ 的平均数是 $\bar{x} + \bar{y}$.

2. 方差的性质

若给定一组数据 x_1, x_2, \dots, x_n 的方差为 s^2 , 则 ax_1, ax_2, \dots, ax_n 的方差为 a^2s^2 , $ax_1 + b, ax_2 + b, \dots, ax_n + b$ 的方差为 a^2s^2 . 特别地, 当 $a=1$ 时, 有 $x_1 + b, x_2 + b, \dots, x_n + b$ 的方差为 s^2 . 这说明将一组数据中的每一个数据都加上一个相同的常数, 方差是不变的, 即不影响数据的波动性.

教材素材变式

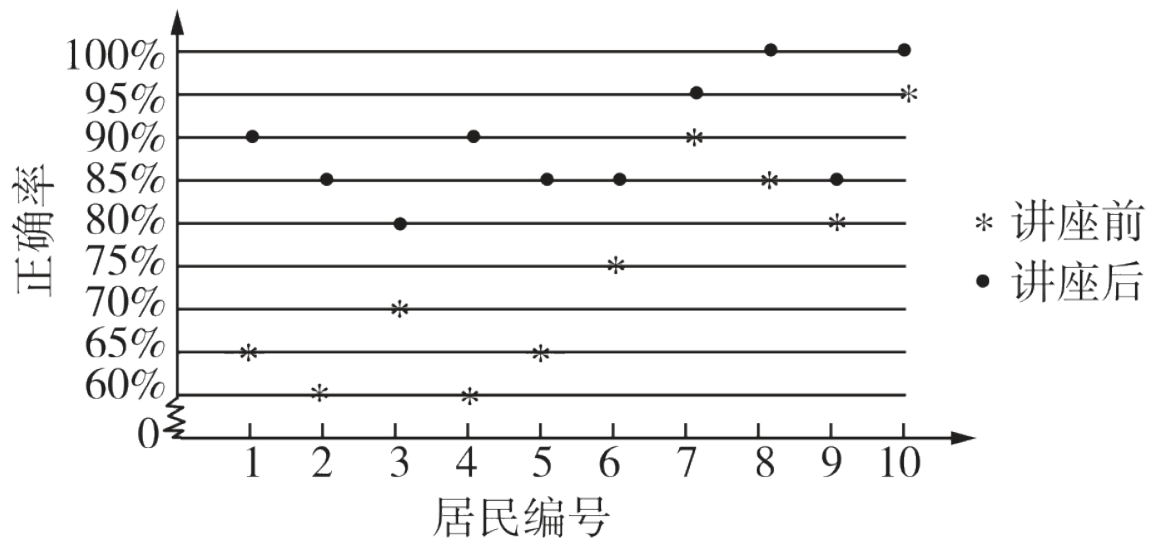
1. [链接人A必修二P205—P213知识] 下列说法正确的是(C)
- A.众数可以准确地反映出总体的情况
 - B.一组数据的平均数一定大于这组数据中的每个数
 - C.平均数、众数与中位数从不同的角度描述了一组数据的集中趋势
 - D.一组数据的方差越大,说明这组数据的波动越小

【解析】众数只利用了出现次数最多的那个值的信息,因此,众数只能传递数据中的信息的很少一部分,并且众数对极端值也不敏感,故A错误.一组数据的平均数不可能大于这组数据的每个数,故B错误.易知C正确.一组数据的方差越小,说明这组数据的波动越小,稳定性越好,一组数据的方差越大,说明这组数据的波动越大,稳定性越差,故D错误.故选C.

2.[人B必修二P79练习B第2题变式，2022全国甲卷（理）]

某社区通过公益讲座以普及社区居民的垃圾分类知识．为了解讲座效果，随机抽取10位社区居民，让他们在讲座前和讲座后各回答一份垃圾分类知识问卷，这10位社区居民在讲座前和讲座后问卷答题的正确率如图，则(**B**)

- A.讲座前问卷答题的正确率的中位数小于70%
- B.讲座后问卷答题的正确率的平均数大于85%
- C.讲座前问卷答题的正确率的标准差小于讲座后正确率的标准差
- D.讲座后问卷答题的正确率的极差大于讲座前正确率的极差



【解析】解法一 根据散点图，讲座前问卷答题正确率的10个数据中，小于或等于70%的为5个，这意味着中位数必然大于70%，故选项A错误。

讲座后问卷答题正确率的10个数据中，有1个为80%，4个为85%，其余数据大于或等于90%。经估算知，平均数大于85%，故选项B正确。

从散点图可以看出，与讲座后问卷答题正确率的数据相比，讲座前问卷答题正确率的数据更为分散，故标准差更大，故选项C错误。

从散点图可直接看出，讲座前问卷答题正确率的极差更大，故选项D错误。故选B。

解法二 根据散点图可知，讲座前问卷答题正确率的中位数为 $\frac{1}{2} \times (70\% + 75\%) > 70\%$ ，故选项A错误。

由散点图中数据计算知，讲座后问卷答题正确率的平均数为89.5%，故选项B正确。

从散点图可以看出，与讲座后问卷答题正确率的数据相比，讲座前问卷答题正确率的数据更为分散，故标准差更大，故选项C错误。

由散点图中数据计算可知，讲座前问卷答题正确率的极差为 $95\% - 60\% = 35\%$ ，讲座后问卷答题正确率的极差为 $100\% - 80\% = 20\%$ ，故选项D错误。故选B。

3. [多选] [人B必修二P91习题5 – 1B第3 (1) 题变式] 下列结论错误的是(**BCD**)

A. 一组数据2, 1, 4, 3, 5, 3的平均数、众数、中位数相同

B. 有A, B, C三种个体按3: 1: 2的比例进行分层抽样调查, 如果抽取的A个体个数为9, 则样本容量为30

C. 若甲组数据的方差为5, 乙组数据为5, 6, 9, 10, 5, 则这两组数据中较稳定的是甲

D. 一组数据1, 2, 2, 2, 3, 3, 3, 4, 5, 6的80%分位数为4

【解析】 A选项：平均数为 $\frac{2+1+4+3+5+3}{6} = 3$ ，众数为3，将数据从小到大依次排列，为1, 2, 3, 3, 4, 5, \therefore 中位数为 $\frac{3+3}{2} = 3$ ，A正确.

B选项：根据样本的抽样比等于各层的抽样比知，样本容量为 $9 \div \frac{3}{3+1+2} = 18$ ，B错误.

C选项：乙组数据的平均数为 $\frac{5+6+9+10+5}{5} = 7$ ，则乙组数据的方差为

$\frac{1}{5} \times [(5-7)^2 + (6-7)^2 + (9-7)^2 + (10-7)^2 + (5-7)^2] = 4.4 < 5$ ， \therefore 这两组数据中较稳定的是乙，C错误.

D选项：该组数据共10个数，由 $10 \times 80\% = 8$ ，知该组数据的80%分位数为 $\frac{4+5}{2} = 4.5$ ，D错误. 故选BCD.

4. [人A必修二P216习题9.2第1题变式] 某校团委为弘扬民族精神，深化爱国主义教育，举办2023年“一二·九”合唱汇演，其中对于高三(1)班的大合唱“保卫黄河”，12位评委的打分情况如下：8.4，9.3，8.9，8.8，8.6，8.2，8.5，8.4，9.2，8.8，8.7，9.4，则这组数据的(D)

A.极差为1

B.众数为8.4

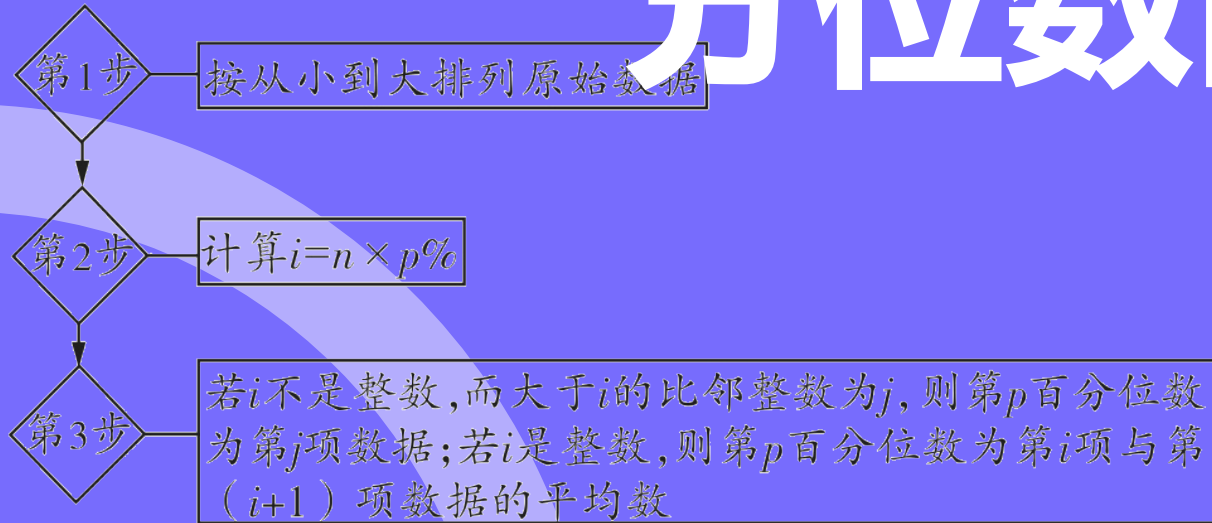
C.80%分位数为8.9

D.第三四分位数为9.05

【解析】将这组数据从小到大依次排列，为8.2，8.4，8.4，8.5，8.6，8.7，8.8，8.8，8.9，9.2，9.3，9.4，(易错对这12个数据一定要在从小到大依次排序的情况下再进行数据分析)则极差为 $9.4 - 8.2 = 1.2$ ，故选项A错误；由于众数为8.4和8.8，故选项B错误；由于 $12 \times 80\% = 9.6$ ，故80%分位数为9.2，故选项C错误；由于 $12 \times 75\% = 9$ ，故第三四分位数为 $\frac{8.9+9.2}{2} = 9.05$ ，故选项D正确。故选D。

知识回顾

计算一组 n 个数据的第 p 百分位数的步骤



常用结论

1.百分位数的特点：（1）一组数据的百分位数可能是这组数据中的数，也可能不是这组数据中的数；（2）第0百分位数为这组数据中的最小的数，第100百分位数为这组数据中的最大的数；（3）一组数据的某些百分位数可能是同一个数.

2.在实际应用中，除了中位数外，常用的分位数还有第25百分位数，第75百分位数.这三个分位数把一组由小到大排列后的数据分成四等份，因此称为四分位数.其中第25百分位数也称为第一四分位数或下四分位数等，第75百分位数也称为第三四分位数或上四分位数等.

5. [苏教必修二P268复习题第11题变式] 某样本中共有5个个体，其值分别为 $a, 0, 1, 2, 3$.若该样本的中位数为1，则实数 a 的取值范围为(**B**)

A.(0,1)

B. $(-\infty, 1]$

C.(1,2)

D.[1,2]

【解析】根据题意，样本数据 $a, 0, 1, 2, 3$ 的中位数为1，知将该样本数据从小到大依次排列后，第三个数据为1，则必有 $a \leq 1$. 故选B.

6. [多选] [人A必修二P209练习第3题变式, 2023新课标 I 卷]有一组样本数据 x_1, x_2, \dots, x_6 , 其中 x_1 是最小值 x_6 是最大值, 则(**BD**)

A. x_2, x_3, x_4, x_5 的平均数等于 x_1, x_2, \dots, x_6 的平均数

B. x_2, x_3, x_4, x_5 的中位数等于 x_1, x_2, \dots, x_6 的中位数

C. x_2, x_3, x_4, x_5 的标准差不小于 x_1, x_2, \dots, x_6 的标准差

D. x_2, x_3, x_4, x_5 的极差不大于 x_1, x_2, \dots, x_6 的极差

【解析】取 $x_1 = 1, x_2 = x_3 = x_4 = x_5 = 2, x_6 = 9$, (指点迷津:适时举特例,可快速排除)则 x_2, x_3, x_4, x_5 的平均数等于2,标准差为0,

x_1, x_2, \dots, x_6 的平均数等于3,标准差为 $\sqrt{\frac{22}{3}} = \frac{\sqrt{66}}{3}$,故A, C均不正确; (技巧:

由于是多选题,故可排除A, C,即选BD)根据中位数的定义,将 x_1, x_2, \dots, x_6 按从小到大的顺序进行排列,中位数是中间两个数的算术平均数,由于 x_1 是最小值, x_6 是最大值,故 x_2, x_3, x_4, x_5 的中位数是将 x_2, x_3, x_4, x_5 按从小到大的顺序排列后中间两个数的算术平均数,与 x_1, x_2, \dots, x_6 的中位数相等,故B正确;根据极差的定义,知 x_2, x_3, x_4, x_5 的极差不大于 x_1, x_2, \dots, x_6 的极差,故D正确.综上,选BD.

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/247104001036010005>