## 基础课51 用样本估计总体

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **考点考向** | **课标要求** | **真题印证** | **考频热度** | **核心素养** |
| 用样本数字特征估计总体集中趋势和离散程度 | 理解 | 2023年新高考Ⅰ卷  2023年新高考Ⅱ卷  2023年全国甲卷  2023年全国乙卷  2023年上海卷 | ★★★ | 数据分析数学运算数学建模 |
| 命题分析预测 | 从近几年高考的情况来看，对用样本估计总体的考查主要是统计图表的应用、总体集中趋势和总体离散程度的估计，单独命题常以客观题的形式出现，也常作为解答题的一问或者一部分出现，属于低、中档题，命题热点是情境化试题.预计2025年的高考命题情况变化不大，本基础课易设置多选题，在备考复习中，要多注意对多选题的训练，做到全面复习 | | | |

### 基础知识·诊断

#### 夯实基础

##### 一、中位数、众数、平均数、百分位数

1.中位数：将一组数据按大小依次排列，处于最中间位置的一个数据（或最中间两个数据的平均数）叫作这组数据的①中位数.

2.众数：在一组数据中出现次数②最多的数据称为这组数据的众数.

3.平均数：一组数据的③算术平均数为这组数据的平均数，个数据,, ,的平均数.

4*.*百分位数

(1)一般地,当总体是连续变量时,给定一个百分数*p*∈(0,1),总体的*p*分位数有这样的特点:总体数据中的任意一个数⑤小于或等于它的可能性是*p.*

(2)计算一组*n*个数据的*p*分位数的一般步骤如下:

第一步,按照从小到大排列原始数据;

第二步,计算*i=*⑥*np*;

第三步,若*i*不是整数,大于*i*的最小整数位*j*,则*p*分位数为第*j*项数据;若*i*是整数,则*p*分位数为第*i*项与第(*i+*1)项数据的平均数*.*

5.四分位数：⑦分位数,分位数,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，因此被称为四分位数.

##### 二、样本的数据特征

如果有个数据,, ,，那么

1.平均数；

2.标准差；

3.方差.

###### 知识 拓展

若数据，， ，的平均数为，则

1.数据，，， ，的平均数是；

2.数据,,, ,的平均数是.

#### 诊断自测

##### 题组1 走出误区

1. 判一判.（对的打“√”,错的打“×”）

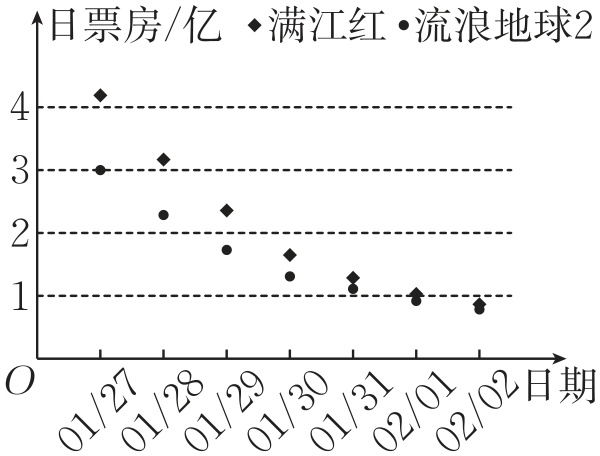
（1） 对一组数据来说，平均数和中位数总是非常接近.( × )

（2） 在频率分布直方图中，最高的小长方形底边中点的横坐标是众数.( √ )

（3） 方差与标准差具有相同的单位.( × )

（4） 若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.( √ )

2. （多选题）（易错题）《满江红》凭借喜剧元素和家国情怀，以25.96亿票房成为档期内票房冠军，另一部科幻续作《流浪地球2》则成为最高口碑电影.如图，这是这两部电影连续7天的日票房情况，则( ABD ).



A. 《满江红》日票房平均数大于《流浪地球2》日票房平均数

B. 《满江红》日票房方差大于《流浪地球2》日票房方差

C. 《满江红》日票房极差小于《流浪地球2》日票房极差

D. 《满江红》日票房的下四分位数小于《流浪地球2》日票房的上四分位数

【**易错点**】不能正确利用数据的波动程度比较方差大小且不清楚四分位数的概念而致误.

[解析]由图表可得《满江红》日票房都大于《流浪地球2》日票房，所以《满江红》日票房平均数大于《流浪地球2》日票房平均数，故正确;由图可得《满江红》日票房单日票房数据波动更大，《满江红》日票房方差大于《流浪地球2》日票房方差，故正确；《满江红》日票房极差大于《流浪地球2》日票房极差，故错误；因为，《满江红》日票房的下四分位数是从小到大排序第2个数，因为，《流浪地球2》日票房的上四分位数是从小到大排序第6个数，《满江红》日票房的下四分位数小于《流浪地球2》日票房的上四分位数，故正确.故选.

##### 题组2 走进教材

3. （人教A版必修改编）四名同学各掷骰子5次，分别记录每次骰子出现的点数.根据四名同学的统计结果，可以判断出一定没有出现点数6的是( C ).

A. 平均数为3，中位数为2 B. 中位数为3，众数为2

C. 平均数为2，方差为2.4 D. 中位数为3，方差为2.8

[解析]对于，当骰子出现结果为1，1，2，5，6时，满足平均数为3，中位数为2，出现了6点，错误；对于，当骰子出现结果为2，2，3，4，6时，满足中位数为3，众数为2，出现了6点，错误；对于，若平均数为2，且出现了6点，则方差，所以一定不会出现6点，正确；对于，当骰子出现结果为1，2，3，3，6时，平均数为3，所以方差为，出现了6点，错误.故选.

4. （双空题）（人教A版必修②P212·例6改编）某学校在对高一年级学生的肺活量进行调查时，采用比例分配的分层随机抽样的方法抽取样本.如果不知道样本数据，只知道抽取了男生20人，他们肺活量的平均数为，方差为10；抽取了女生30人，她们肺活量的平均数为，方差为20.由此可得抽取的高一年级学生肺活量的平均数为2700，方差为60016.

[解析]把抽取的男生肺活量数据的平均数记为,方差记为,女生肺活量数据的平均数记为,方差记为；把总样本数据的平均数记为,方差记为.由,,根据比例分配的分层随机抽样中总样本平均数与各层样本平均数的关系，可得总样本数据的平均数，则.由此可得抽取的高一年级学生肺活量的平均数为,方差为60016.

##### 题组3 走向高考

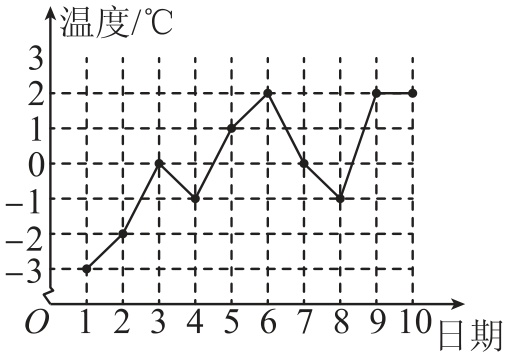
5. [2023·上海卷改编]现有某地一年四个季度的（单元：亿元），第一季度的为232亿元，第四季度的为241亿元，四个季度的逐季度增长，且中位数与平均数相同，则该地一年的为946亿元.

[解析]设第二季度为亿元，第三季度为亿元，则，因为中位数与平均数相同，所以，则，所以该地一年的为（亿元）.

### 考点聚焦·突破

#### 考点一 总体百分位数的估计［自主练透］

1. [2024·安徽模拟]如图，这是根据某市3月1日至3月10日的最低气温（单位：）的情况绘制的折线统计图，由图可知这10天最低气温的第80百分位数是( D ).



A. B. 0 C. 1 D. 2

[解析]由折线图可知，这10天的最低气温按照从小到大的顺序排列为，，，，0，0，1，2，2，2，因为共有10个数据，所以，是整数，则这10天最低气温的第80百分位数是.故选.

2. [2024·浙江模拟]（多选题）已知100个数据的第75百分位数是，则下列说法不正确的是( ABD ).

A. 这100个数据中一定有75个数小于或等于9.3

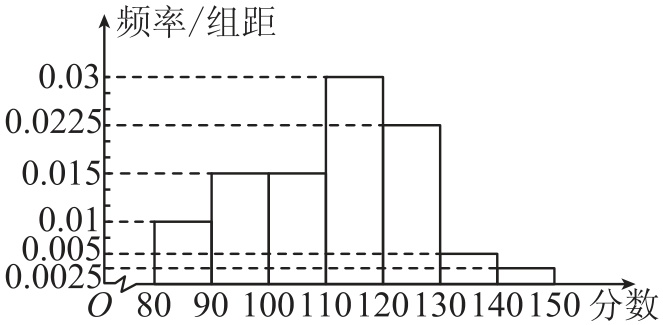
B. 把这100个数据从小到大排列后，是第75个数据

C. 把这100个数据从小到大排列后，是第75个数据和第76个数据的平均数

D. 把这100个数据从小到大排列后，是第75个数据和第74个数据的平均数

[解析]因为为整数，所以把这100个数据从小到大排列后，第75个数据和第76个数据的平均数为第75百分位数，是，则正确，其他选项均不正确.故选.

3. 若将高三某班60名学生参加某次数学模拟考试所得的成绩（成绩均为整数）整理后画出频率分布直方图如图所示，则该班的模拟考试成绩的分位数约为124.44.（结果保留两位小数）



[解析]由频率分布直方图可知，分数在120分以下的学生所占的比例为，分数在130分以下的学生所占的比例为，

因此，分位数一定位于内.

因为，

## 所以该班的模拟考试成绩的分位数约为124.44.

4. （双空题）一个容量为20的样本，若其数据按从小到大的顺序排列为1，2，2，3，5，6，6，7，8，8，9，10，13，13，14，15，17，17，18，18，则该组数据的第75百分位数为14.5，第86百分位数为17.

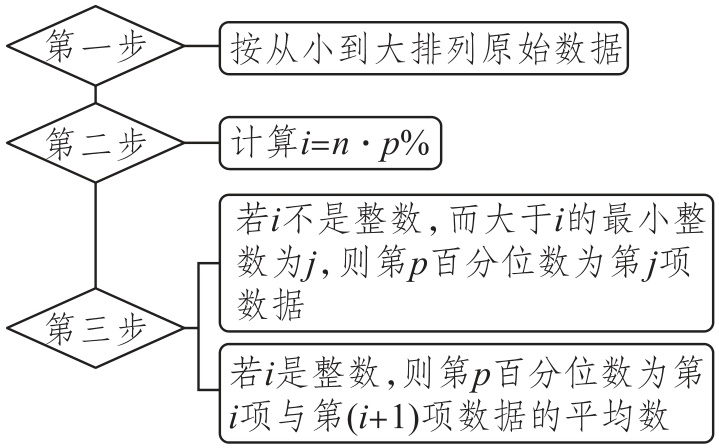
[解析]，

第75百分位数为.

， 第86百分位数为第18个数据17.



**计算一组数据的第百分位数的步骤**



#### 考点二 总体集中趋势的估计［多维探究］

##### 样本的数字特征角度1

## 典例1 已知甲、乙两班各50人，某次数学考试的成绩情况如表所示：

|  |  |  |
| --- | --- | --- |
| 分数段 | 甲班人数 | 乙班人数 |
|  | 1 | 3 |
|  | 3 | 2 |
|  | 9 | 5 |
|  | 16 | 14 |
|  | 10 | 11 |
|  | 7 | 8 |
|  | 4 | 7 |

各分数段成绩视为均匀分布，有以下结论：①甲班平均成绩低于乙班；②甲班成绩的中位数与乙班相同；③甲班成绩的方差比乙班成绩的方差小.其中正确结论的序号是( B ).

A. ① B. ①③ C. ②③ D. ①②③

[解析]由题可知，，

，所以甲班的平均成绩低于乙班的平均成绩，故①正确；

因为甲班的中位数位于分数段，乙班的中位数位于分数段，所以甲班的中位数低于乙班的中位数，故②错误；

因为，

，所以甲班成绩的方差比乙班成绩的方差小，故③正确.

综上，正确结论的序号是①③.故选.



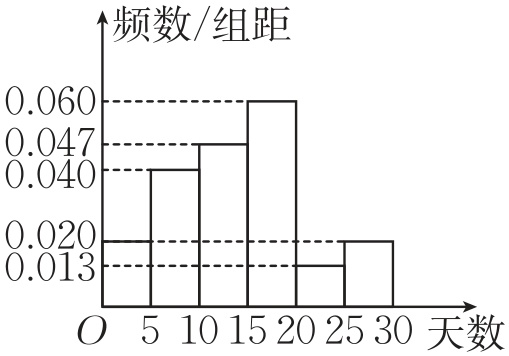
1.求平均数时要注意数据的个数，不要重计或漏计.

2.求中位数时一定要先对数据按大小排序，若最中间有两个数据，则中位数是这两个数据的平均数.

3.若有两个或两个以上的数据出现得最多，且出现的次数一样，则这些数据都叫众数；若一组数据中每个数据出现的次数一样多，则没有众数.

##### 频率分布直方图中的数字特征角度2

典例2 某滑冰馆统计了2023年9月1日到30日某小区居民在该滑冰馆的锻炼天数，得到如图所示的频率分布直方图（将频率视为概率），则下列说法正确的是( C ).



A. 该小区居民在该滑冰馆的锻炼天数在区间内的最少

B. 估计该小区居民在该滑冰馆的锻炼天数的中位数为16

C. 估计该小区居民在该滑冰馆的锻炼天数的平均值大于14

D. 估计该小区居民在该滑冰馆的锻炼天数超过15天的概率为0.456

[解析]频率分布直方图中，面积最小的矩形条所在的区间为，即样本中区间内的数据频率最小，频数也最小，故错误；

由频率分布直方图可得，前三个小矩形的面积之和为，所以估计该小区居民在该滑冰馆的锻炼天数的中位数小于15，故错误；

由频率分布直方图可得，，故正确；

由频率分布直方图可得，该小区居民在该滑冰馆的锻炼天数超过15天的频率为，故锻炼天数超过15天的概率为，故错误.故选.



**频率分布直方图的数字特征**

1.众数：众数一般用频率分布表中频率最高的一组的组中值来表示，即在样本数据的频率分布直方图中，最高小长方形的底边中点的横坐标.

2.中位数：在频率分布直方图中，中位数左边和右边的直方图的面积应该相等.

3.平均数：平均数在频率分布直方图中等于组中值与对应频率之积的和.

##### 多维训练

1. [2023·新高考Ⅰ卷]（多选题）有一组样本数据,, ,，其中是最小值，是最大值，则( BD ).

A. ,,,的平均数等于,, ,的平均数

B. ,,,的中位数等于,, ,的中位数

C. ,,,的标准差不小于,, ,的标准差

D. ,,,的极差不大于,, ,的极差

[解析]对于，设,,,的平均数为，,, ,的平均数为，

则，

因为没有确定,的大小关系，所以无法判断,的大小，故错误.

对于，不妨设，可知,,,的中位数与,, ,的中位数均为，故正确.

对于，因为是最小值，是最大值，所以,,,的波动性不大于,, ,的波动性，即,,,的标准差不大于,, ,的标准差，故错误.

对于，不妨设，则，当且仅当,时，等号成立，故正确.故选.

2. 为普及科学用眼知识，提高群众健康水平，预防眼疾，某区残联在残疾人综合服务中心开展“全国爱眼日”有奖答题竞赛活动.已知5位评委老师按百分制（只打整数分）分别对某参赛小队评分，可以判断出一定有评委打满分的是( D ).

A. 平均数为98，中位数为98 B. 中位数为96，众数为99

C. 中位数为97，极差为9 D. 平均数为98，极差为6

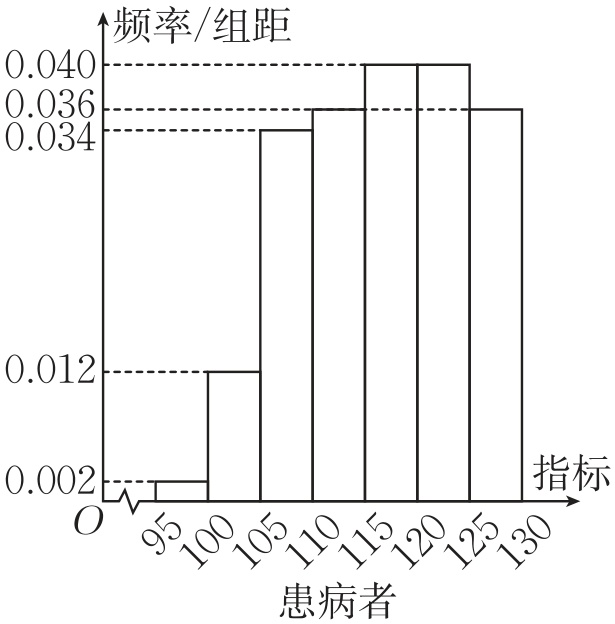
[解析]对于，当打分结果为98,98,98,98,98时，满足平均数为98，中位数为98，所以错误；

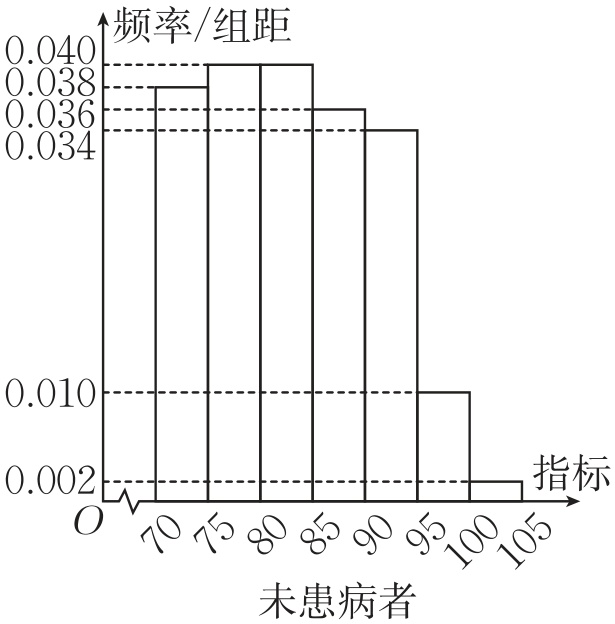
对于，当打分结果为99,99,96,95,94时，满足中位数为96，众数为99，所以错误；

对于，当打分结果为89,97,97,97,98时，满足中位数为97，极差为9，所以错误；

对于，假设没有评委打满分，结合极差为6可得总成绩，则平均数，与选项不符，故假设不成立，所以当平均数为98，极差为6时，一定有评委打满分，所以正确.故选.

3. [2023·新高考Ⅱ卷节选]（双空题）某研究小组经过研究发现某种疾病的患病者与未患病者的某项医学指标有明显差异，经过大量调查，得到如图所示的患病者和未患病者该指标的频率分布直方图：





利用该指标制定一个检测标准，需要确定临界值，将该指标大于的人判定为阳性，小于或等于的人判定为阴性.此检测标准的漏诊率是将患病者判定为阴性的概率，记为；误诊率是将未患病者判定为阳性的概率，记为.假设数据在组内均匀分布，以事件发生的频率作为相应事件发生的概率.当漏诊率时，临界值97.5，误诊率.

[解析]依题可知，患病者该指标在内的频率为，所以，

所以，解得，

.

#### 考点三 总体离散程度的估计［师生共研］

典例3 [2023·全国乙卷]某厂为比较甲、乙两种工艺对橡胶产品伸缩率的处理效应，进行了10次配对试验，每次配对试验选用材质相同的两个橡胶产品，随机地选其中一个用甲工艺处理，另一个用乙工艺处理，测量处理后的橡胶产品的伸缩率.甲、乙两种工艺处理后的橡胶产品的伸缩率分别记为，.试验结果如表所示.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 试验序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 伸缩率 | 545 | 533 | 551 | 522 | 575 | 544 | 541 | 568 | 596 | 548 |
| 伸缩率 | 536 | 527 | 543 | 530 | 560 | 533 | 522 | 550 | 576 | 536 |

记，记,, ,的样本平均数为，样本方差为.

（1）求，；

（2）判断甲工艺处理后的橡胶产品的伸缩率较乙工艺处理后的橡胶产品的伸缩率是否有显著提高.（如果，则认为甲工艺处理后的橡胶产品的伸缩率较乙工艺处理后的橡胶产品的伸缩率有显著提高，否则不认为有显著提高）

[解析]（1），

，

，

的值分别为9,6,8,,15,11,19,18,20,12，

故.

（2）由（1）知，，故有,

所以认为甲工艺处理后的橡胶产品的伸缩率较乙工艺处理后的橡胶产品的伸缩率有显著提高.



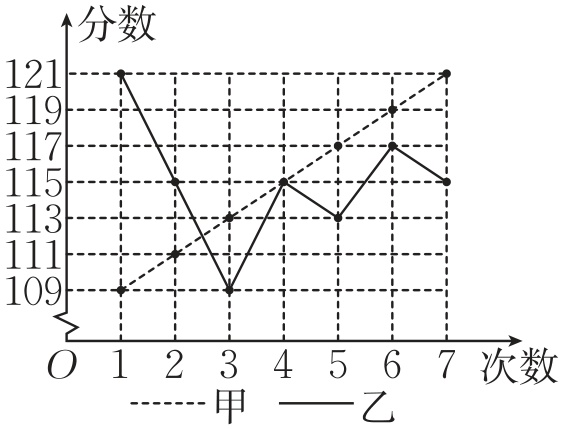
**利用样本的数字特征解决优化问题的依据**

1.平均数反映了数据取值的平均水平，标准差、方差描述了一组数据围绕平均数波动的大小.标准差、方差越大，数据的离散程度就越大，越不稳定；标准差、方差越小，数据的离散程度就越小，越稳定.

2.用样本估计总体就是利用样本的数字特征来描述总体的数字特征.

##### 针对训练

某高校数学组积极研讨教学策略,先采取甲、乙两套方案教学，并对分别采取两套方案教学的班级的7次测试成绩进行统计，统计结果如图所示：



（1） 请填写下表.（要求写出计算过程）

|  |  |  |
| --- | --- | --- |
|  | 平均数 | 方差 |
| 甲 |  |  |
| 乙 |  |  |

[解析]由图象中的数据可得，

，

，.

（2） 从下列三个不同的角度对这次方案选择的结果进行分析：

①从平均数和方差相结合看（分析哪种方案的成绩更好）；

②从折线图上两种方案的成绩走势看（分析哪种方案更有潜力）.

[解析]①甲、乙两种方案的成绩的平均数相等，且，故乙方案更稳定；

②从折线图的走势上看甲更好，使用甲方案的成绩稳步提高，而使用乙方案的成绩不稳定，忽上忽下，故甲方案更有潜力.