**14.4 用样本估计总体**

**【考点梳理】**

**考点一、平均数、中位数、众数的应用**

**考点二、极差、方差、标准差的计算与应用**

**考点三：各数据加减乘除对方差、平均数的影响**

**考点四：频率分布直方图中的方差、标准差**

**考点五：百分位数**

**考点六：用样本估计总体的综合问题**

**【知识梳理】**

**知识点一、用样本估计总体的集中趋势参数**

1.平均数

(1)均值：一般地、把总体中所有数据的算术平均数称为总体的均值，它通常可以代表总体的水平。

(2)平均数：如果给定的一组数是，，…，是则这组数的平均数为 即 。

(3)一般地，若取值为，，…，的频率分别为，，…，则其平均数为+++……+。

(4)如果将总体分为k层，第j层抽取的样本为，，……，第j层的样本量为，样本平均数为，j=1，2，…，k。记，则所有数据的样本平均数为。

2.众数

一般地，我们将一组数据中出现次数最多的那个数据叫作该组数据的众数。众数是一种刻画数据集中趋势的度量值。

3.中位数

一般地，将一组数据按照从小到大的顺序排成一列，如果数据的个数为奇数，那么排在正中间的数据就是这组数据的中位数；如果数据的个数为偶数，那么，排在正中间的两个数据的平均数即为这组数据的中位数。

**技巧归纳：平均数、中位数和众数的特点**

(1)平均数的大小与一组数据里每一个数据均有关系，任何一个数据的变化都会引起平均数的变化.容易受极端值的影响；

(2)中位数仅与数据排列位置有关，某些数据的变动对中位数没有影响，中位数可能在所给数据中（数据个数为奇数或数据个数为偶数且中间两数相等），也可能不在所给的数据中（数据个数为偶数且中间两数不相等）。当一组数据的个别数据变动较大时，可以用中位数描述其集中趋势

(3)众数只与这组数据中的部分数据有关，当一组数据中有多个数据多次重复出现时，众数往往更能反映问题。

4.频率直方图中的数字特征

(1)样本平均数可以用每个小矩形底边中点的横坐标与小矩形的面积的乘积之和近似代替。

(2)样本中位数可依据“中位数左边和右边的直方图的面积相等”来求出。

(3)样本众数可以用最高的矩形底边中点的横坐标近似代替。

**知识点二、用样本估计总体的离散程度参数**

1.极差我们把一组数据的最大值与最小值的差称为极差。

2.方差和标准差

假设一组数据是，，…，用表示这组数据的平均数，则为这组数据的方差。有时为了计算方差的方便，我们还把方差写成的形式。我们对方差开平方，取它的算

术平方根，即为这组数据的标准差。

3.总体方差(标准差)和样本方差(标准差)

①如果总体中所有个体的变量值分别为，，…，，总体平均数为，则称

为总体方差， 为总体标准差。

②如果总体的N个变量值中，不同的值共有k(k≤N)个，不妨记为，，…，，其中出现的频数为(i=1，2，…，k)，

则总体方差为 。

如果一个样本中个体的变量值分别为，，…，，样本平均数为，则称为样本方差， 为样本标准差。

3.标准差

因为方差与原始数据的单位不同，且平方后可能夸大了离差的程度，所以我们将方差的算术平方根称为这组数据的标准差.标准差刻画了数据的离散程度或波动幅度，标准差越大，数据的离散程度越大；标准差越小，数据的离散程度越小。

4.一般地，若取值为，，…，的频率分别为，，…，，则其方差为 。

5.分层抽样数据的方差

一般地，如果总体分为k层，第j层抽取的样本为，，……，第j层的样本量为，样本平均数为，样本方差为，=1,2，……，k。记。

**知识点三、百分位数**

1.k百分位数的特点

一般地，一组数据的k百分位数是这样一个值，它使得这组数据中至少有k%的数据小于或等于，且至少有(100-k)%的数据大于或等于

2.计算有n个数据的大样本的k百分位数的一般步骤

第1步，将所有数值按从小到大的顺序排列；

第2步，计算

第3步，如果结果为整数，那么k百分位数位于第k· 位和下一位数之间，通常取这两个位置上数值的平均数为k百分位数；

第4步，如果 不是整数，那么将其向上取整(即其整数部分加上1)，在该位置上的数值即为k百分位数。

3.四分位数常用的分位数有25百分位数，50百分位数(即中位数)和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25百分位数称为下四分位数，75百分位数称为上四分位数。

**【题型归纳】**

**题型一、平均数、中位数、众数的应用**

1．（20-21高一下·河北·期末）某校举行校园歌手大赛，6位评委对某选手的评分分别为9.2，9.5，8.8，9.9，8.9，9.5，设该选手得分的平均数为*x*，中位数为*y*，众数为*z*，则（ ）

A． B． C． D．

2．（2024高一上·湖南邵阳·竞赛）如果将一组数据5、4、6、5、4、13、5依次重复写10次，会得到70个数组成的一组新数据，关于这组新数据的中位数、众数、平均数，下列说法正确的是（ ）

A．中位数和众数都是5 B．众数是10

C．中位数是4 D．中位数、平均数都是5

3．（23-24高二上·四川成都·期中）已知甲、乙两支篮球队各6名队员某场比赛的得分数据（单位：分）从小到大排列为如下：甲队：7，12，12，20，，31；乙队：8，9，19，，25，28.这两组数据的中位数相等，且平均值也相等，则*x*和*y*的值分别为（ ）

A．2和3 B．0和2 C．0和3 D．2和4

**题型二、极差、方差、标准差的计算与应用**

4．（23-24高一下·浙江宁波·期中）甲乙丙丁四位同学各掷5次骰子并记录点数，方差最大的是（ ）

甲：4 5 4 5 5 乙：4 2 3 4 3

丙：2 3 2 3 4 丁：6 1 2 6 1

A．甲 B．乙 C．丙 D．丁

5．（2024·全国·模拟预测）已知总体划分为3层，通过分层随机抽样，各层抽取的样本容量分别为，，，样本平均数分别为，，，样本方差分别为，，，若，则（ ）

A．

B．

C．总体样本平均数

D．当时，总体方差

6．（2024·山西朔州·一模）已知一组样本数据分别为，若这组数据的平均数为4，则数据，的方差为（ ）

A． B．8 C．16 D．24

**题型三：各数据加减乘除对方差、平均数的影响**

7．（2024·湖南长沙·二模）已知样本数据的平均数和标准差均为4，则数据的平均数与方差分别为（ ）

A． B． C． D．

8．（23-24高一下·浙江绍兴·期中）已知样本数据的平均数和标准差均为4，则数据的平均数与标准差分别为（ ）

A． B． C． D．

9．（2023·全国·模拟预测）已知样本数据，，…，的平均数为，方差为，若样本数据，，…，的平均数为，方差为，则（ ）

A．5 B． C．1或5 D．或

**题型四：频率分布直方图中的方差、标准差**

10．（2024·陕西西安·模拟预测）某校为了解在校学生对中国传统文化的传承认知情况，随机抽取了100名学生进行中国传统文化知识考试，并将这100名学生成绩整理得到如下频率分布直方图．根据此频率分布直方图（分成，，，，，六组），下列结论中不正确的是（ ）



A．图中的

B．若从成绩在，，内的学生中采用分层抽样抽取10名学生，则成绩在内的有3人

C．这100名学生成绩的中位数约为65

D．若同一组中的数据用该组区间的中点值作代表，则这100名学生的平均成绩约为68.2

11．（22-23高一下·吉林长春·期末）某校1500名学生参加交通安全知识竞赛，随机抽取了100名学生的竞赛成绩（单位：分），成绩的频率分布直方图如图所示，则下列说法正确的是（ ）

A．频率分布直方图中的值为0.0045

B．估计这100名学生竞赛成绩的中位数为73

C．估计这100名学生竞赛成绩的众数为80

D．估计总体中成绩落在内的学生人数为525

12．（2023·天津河北·二模）某校举行知识竞赛，对全校参赛的1000名学生的得分情况进行了统计，把得分数据按，，，，分成5组，得到如图所示的频率分布直方图，则下列说法不正确的是（ ）



A．图中的*x*值为0.020 B．得分在的人数为400

C．这组数据的极差为50 D．这组数据的平均数的估计值为77

**题型五：百分位数**

13．（2024高一下·全国·专题练习）小明同学统计了他最近10次的数学考试成绩，得到的数据分别为92，85，87，91，95，90，88，83，98，96．则这组数据的分位数是（ ）

A．92 B．91.5

C．91 D．90

14．（2024·天津·二模）某校举办了数学知识竞赛，把1000名学生的竞赛成绩（满分100分，成绩取整数）按，，，分成四组，并整理成如图所示的频率分布直方图，则下列说法正确的为（ ）



A．的值为0.015 B．估计这组数据的众数为80

C．估计这组数据的第60百分位数为87 D．估计成绩低于80分的有350人

15．（2024·安徽安庆·二模）在一次学科核心素养能力测试活动中，随机抽取了100名同学的成绩（评分满分为100分），将所有数据按，，，，，进行分组，整理得到频率分布直方图如图所示，则估计这次调查数据的第64百分位数为（ ）



A．80 B．78 C．76 D．74

**题型六：用样本估计总体的综合问题**

16．（2024高一下·全国·专题练习）为了解学生的周末学习时间（单位：小时），高一年级某班班主任对本班40名学生某周末的学习时间进行了调查，将所得数据整理绘制出如图所示的频率分布直方图，根据直方图所提供的信息：

(1)求该班学生周末的学习时间不少于20小时的人数；

(2)①估计这40名同学周末学习时间的25%分位数；

②将该班学生周末学习时间从低到高排列，那么估计第10名同学的学习时长；

(3)如果用该班学生周末的学习时间作为样本去推断该校高一年级全体学生周末的学习时间，这样推断是否合理？说明理由．

17．（23-24高一下·福建泉州·期中）果切是一种新型水果售卖方式，商家通过对整果进行消洗、去皮、去核、冷藏等操作后，包装组合销售，在“健康消费”与“瘦身热潮”的驱动下，果切更能满足消费者的即食需求.

(1)统计得到10名中国果切消费者每周购买果切的次数依次为：1，7，4，7，4，6，6，3，7，5，求这10个数据的平均数与方差；

(2)统计600名中国果切消费者的年龄，他们的年龄均在5岁到55岁之间，按照，，，，分组，得到频率分布直方图.

①估计这600名中国果切消费者中年龄不小于35岁的人数；

②估计这600名中国果切消费者年龄的中位数及平均数（结果保留整数）.

18．（2024高一下·江苏·专题练习）“一带一路”是“丝绸之路经济带”和“21世纪海上丝绸之路”的简称.某市为了了解人们对“一带一路”的认知程度，对不同年龄和不同职业的人举办了一次“一带一路”知识竞赛，满分100分（90分及以上为认知程度高）.现从参赛者中抽取了*x*人，按年龄分成5组，第一组：，第二组：，第三组：，第四组：，第五组：，得到如图所示的频率直方图，已知第一组有6人.



(1)求*x*；

(2)从该市大学生、军人、医务人员、工人、个体户五种人中用分层抽样的方法依次抽取6人，42人，36人，24人，12人，分别记为1～5组，从这5个按年龄分的组和5个按职业分的组中每组各选派1人参加知识竞赛，分别代表相应组的成绩，年龄组中1～5组的成绩分别为93，96，97，94，90，职业组中1～5组的成绩分别为93，98，94，95，90.

①分别求5个年龄组和5个职业组成绩的平均数和方差；

②以上述数据为依据，评价5个年龄组和5个职业组对“一带一路”的认知程度.

**【双基达标】**

**一、单选题**

19．（2024·江西·二模）从甲队60人、乙队40人中，按照分层抽样的方法从两队共抽取10人，进行一轮答题．相关统计情况如下：甲队答对题目的平均数为1，方差为1；乙队答对题目的平均数为1.5，方差为0.4，则这10人答对题目的方差为（ ）

A．0.8 B．0.675 C．0.74 D．0.82

20．（2024高一下·全国·专题练习）已知一组数据从小到大排列为0，4，5，6，8，10，12，15，则这组数据的分位数、分位数分别是（ ）

A．5.5,10 B．5.5,12 C．6,11 D．6,10

21．（2024高三·全国·专题练习）为了分析某次数学考试的情况，随机抽取了若干学生，将其考试成绩（单位：分）分组为[60，70），[70，80），[80，90），[90，100），[100，110），[110，120），[120，130），[130，140），[140，150]，并绘制成如图所示的频率分布直方图，据此可估计该次考试成绩的中位数为*m*，则*m*的值为（ ）

A．99.8 B．100.8 C．101.8 D．102.8

22．（23-24高一下·甘肃·期中）某高校为宣扬中华文化，举办了“论语吟唱”的比赛，在比赛中，由A，B两个评委小组（各9人）给参赛选手打分．根据两个评委小组对同一名选手的打分绘制成如图所示折线图，则下列说法正确的是（ ）

A．B组打分的极差小于A组打分的极差 B．B组打分的中位数为75

C．A组的意见相对一致 D．A组打分的众数为50

23．（23-24高三下·四川泸州·阶段练习）已知某人收集一个样本容量为50的一组数据，并求得其平均数为70，方差为75，现发现在收集这些数据时，其中两个数据记录有误，一个错将80记录为60，另一个错将70记录为90，在对错误数据进行更正后，重新求得样本的平均数为，方差为，则（ ）

A． B．

C． D．

24．（23-24高一下·甘肃金昌·阶段练习）某校高一年级15个班参加朗诵比赛的得分如下（单位：分）：

则这组数据的分位数为（ ）

A．91 B．90 C．89.5 D．89

25．（2024高一下·全国·专题练习）从某企业生产的某种产品中抽取100件，测量这些产品的一项质量指标值，由测量结果得到如下频数分布表：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 质量指标值分组 |  |  |   |   |  |
| 频数 | 6 | 26 | 38 | 22 | 8 |

(1)作出这些数据的频率分布直方图；

(2)估计这种产品质量指标值的平均数及方差（同一组中的数据用该组区间的中点值作代表）；

(3)根据以上抽样调查数据，能否认为该企业生产的这种产品符合“质量指标值不低于95的产品至少要占全部产品的”的规定？

26．（23-24高二下·四川绵阳·阶段练习）文明城市是反映城市整体文明水平的综合性荣誉称号，作为普通市民，既是文明城市的最大受益者，更是文明城市的主要创造者，某市为提高市民对文明城市创建的认识，举办了“创建文明城市”知识竞赛，从所有答卷中随机抽取100份作为样本，将样本的成绩（满分100分，成绩均为不低于40分的整数）分成六段：，，得到如图所示的频率分布直方图.



(1)求频率分布直方图中的值；

(2)求样本成绩的第75百分位数；

(3)现从该样本成绩在与两个分数段内的市民中按分层抽样选取6人，求从这6人中随机选取2人，且2人的竞赛成绩之差的绝对值大于20的概率.

**【高分突破】**

**一、单选题**

27．（2024·江苏南通·二模）一组样本数据删除一个数后，得到一组新数据：10，21，25，35，36，40.若这两组数据的中位数相等，则删除的数为（ ）

A．25 B．30 C．35 D．40

28．（2024·内蒙古呼和浩特·一模）已知样本数据的平均数为､方差为，若样本数据，的平均数为，方差为，则（ ）

A． B． C． D．

29．（2024·福建·模拟预测）某单位共有*A*、*B*两部门，1月份进行服务满意度问卷调查，得到两部门服务满意度得分的频率分布条形图如下.设*A*、*B*两部门的服务满意度得分的第75百分位数分别为，，方差分别为，，则（ ）



A．， B．，

C．， D．，

30．（2024·江西赣州·一模）若一组样本数据的方差为，则样本数据的方差为（ ）

A．1 B．2 C．2.5 D．

**二、多选题**

31．（2024高一下·全国·专题练习）下列说法中正确的是（ ）

A．数据2，4，6，8的中位数是4，6

B．数据1，2，2，3，4，4的众数是2，4

C．一组数据的平均数、众数、中位数有可能是同一个数

D．8个数据的平均数为5，另3个数据的平均数为7，则这11个数据的平均数是

32．（2024高一下·全国·专题练习）已知100个数据的75%分位数是9.3，则下列说法正确的是（ ）

A．这100个数据中至少有75个数小于或等于9.3

B．把这100个数据从小到大排列后，9.3是第75个数据

C．把这100个数据从小到大排列后，9.3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，9.3是第74个数据和第75个数据的平均数

33．（2024·全国·模拟预测）在某市高三年级等行的一次数学期末考试中，为了解考生的成绩情况，随机抽取了50名考生的成绩，作出的频率分布直方图如图，成绩排在前的学生将获得“优秀学生”称号，则（ ）



A．估计该市考生的成绩低于60分的比例为

B．估计该市考生成绩的众数为60

C．估计该市考生成绩的平均数为70.6

D．估计该市82分以上的考生将获得“优秀学生”称号

34．（23-24高一下·全国·课后作业）【多选】一组数据的平均值为5，方差为2，极差为7，中位数为6，记，，，，的平均值为，方差为，极差为，中位数为，则（ ）

A． B． C． D．

35．（2024·广东汕头·一模）某次数学考试后，为分析学生的学习情况，某校从某年级中随机抽取了名学生的成绩，整理得到如图所示的频率分布直方图.为进一步分析高分学生的成绩分布情况，计算得到这名学生中，成绩位于内的学生成绩方差为，成绩位于内的同学成绩方差为.则（ ）

参考公式：样本划分为层，各层的容量､平均数和方差分别为：、、；、、.记样本平均数为，样本方差为，.



A．

B．估计该年级学生成绩的中位数约为

C．估计该年级成绩在分及以上的学生成绩的平均数为

D．估计该年级成绩在分及以上的学生成绩的方差为

**三、填空题**

36．（2024高一下·全国·专题练习）从高三抽出50名学生参加数学竞赛，由成绩得到如图的频率分布直方图，则估计这50名学生成绩的分位数为 分．



37．（2024高一下·全国·专题练习）某校从高一年级中随机抽取部分学生，将他们的期末数学测试成绩分成6组：加以统计，得到如图所示的频率分布直方图．据此统计，期末数学测试成绩不少于分位数的分数至少为 分．



38．（2024·全国·模拟预测）某小组5位同学各拋掷一枚正方体骰子，将正面向上的点数按从小到大的顺序记录下来，得到一组统计数据．已知这组数据的平均数为整数，最大值为6，中位数为3，方差为1.6，则这组数据的众数为 ．

39．（23-24高三上·上海宝山·期末）在某次比赛中运动员五轮的成绩互不相等，记为，平均数为，若随机删去其中一轮的成绩，得到一组新数据，记为，平均数为，下面说法正确的是 .（写出所有正确选项）

①新数据的极差可能等于原数据的极差.

②新数据的中位数可能等于原数据的中位数.

③若，则新数据的方差一定大于原数据方差.

④若，则新数据的第40百分位数一定大于原数据的第40百分位数.

**四、解答题**

40．（2024高二下·湖南娄底·学业考试）2023年10月22日，汉江生态城2023襄阳马拉松在湖北省襄阳市成功举行，志愿者的服务工作是马拉松成功举办的重要保障，襄阳市新时代文明实践中心承办了志愿者选拔的面试工作．现随机抽取了100名候选者的面试成绩，并分成五组：第一组，第二组，第三组，第四组，第五组，绘制成如图所示的频率分布直方图．已知第一、二组的频率之和为0.3，第一组和第五组的频率相同．



(1)估计这100名候选者面试成绩的平均数和第25百分位数；

(2)现从以上各组中用分层随机抽样的方法选取20人，担任本市的宣传者．若本市宣传者中第二组面试者的面试成绩的平均数和方差分别为62和40，第四组面试者的面试成绩的平均数和方差分别为80和70，据此估计这次第二组和第四组所有面试者的面试成绩的方差．

41．（2024高一下·全国·专题练习）文明城市是反映城市整体文明水平的综合性荣誉称号，作为普通市民，既是文明城市的最大受益者，更是文明城市的主要创造者.某市为提高市民对文明城市创建的认识，举办了“创建文明城市”知识竞赛，从所有答卷中随机抽取100份作为样本，将样本的成绩（满分100分，成绩均为不低于40分的整数）分成六段：，，，得到如图所示的频率分布直方图.



(1)求频率分布直方图中的值；

(2)求样本成绩的第75百分位数；

(3)已知落在的平均成绩是56，方差是7，落在的平均成绩为65，方差是4，求两组成绩的总平均数和总方差.

