

第十章

统计与统计案例

第一节 统计

本节主要包括 2 个知识点：□

1. 随机抽样；□

2. 用样本估计总体.

突破点(一) 随机抽样

突破点(二) 用样本预计总体

03

全国卷5年真题集中演练——明规律

课时达标检测



01

突破点(一) 随机抽样



[基本知识]

1. 简单随机抽样

(1)定义：设一个总体含有 N 个个体，从中逐个不放回地抽取 n 个个体作为样本($n \leq N$)，如果每次抽取时总体内的各个个体被抽到的机会都相等，就把这种抽样方法叫做简单随机抽样。

(2)最常用的简单随机抽样的方法：抽签法和随机数法。

2. 系统抽样

在抽样时，将总体分成均衡的几个部分，然后按照事先确定的规则，从每一部分抽取一个个体，得到所需要的样本，这种抽样方法叫做系统抽样(也称为机械抽样)。



3. 分层抽样

在抽样时，将总体分成互不交叉的层，然后按照一定的比例，从各层独立地抽取一定数量的个体，将各层取出的个体合在一起作为样本，这种抽样方法是一种分层抽样。



4. 三种抽样方法的比较

类别	共同点	各自特点	• 相互联络	适用范围
简单随机抽样	<ul style="list-style-type: none"> • 均为不放回抽样，且抽样过程中每个个体被抽 	<ul style="list-style-type: none"> • 从总体中逐一抽取 	<ul style="list-style-type: none"> • 是后两种方法基础 	<ul style="list-style-type: none"> • 总体中个数较少
系统抽样		<ul style="list-style-type: none"> • 将总体均分成几部分，按事先确定规则在各部分中抽取 	<ul style="list-style-type: none"> • 在起始部分抽取简单随机抽样 	<ul style="list-style-type: none"> • 元素个数很多且均衡总体抽样
		<ul style="list-style-type: none"> • 将总体 	<ul style="list-style-type: none"> • 各自比 	



[基本能力]

1. 判断题

(1)简单随机抽样是一种不放回抽样。 (✓)

(2)简单随机抽样每个个体被抽到的机会不一样，与先后有关。 (×)

(3)系统抽样在起始部分抽样时采用简单随机抽样。 (✓)

(4)要从 1 002 个学生中用系统抽样的方法选取一个容量为 20 的样本，需要剔除 2 个学生，这样对被剔除者不公平。 (×)

(5)分层抽样中，每个个体被抽到的可能性与层数及分层有关。 (×)



2. 填空题

(1)利用简单随机抽样从含有 8 个个体的总体中抽取一个容量为 4 的样本，则总体中每个个体被抽到的概率是_____.

解析：总体个数为 $N=8$ ，样本容量为 $M=4$ ，则每一个个体被

抽到的概率为 $P=\frac{M}{N}=\frac{4}{8}=\frac{1}{2}$.

答案： $\frac{1}{2}$



(2) 老师在班级 50 名学生中，依次抽取学号为 5,10,15,20,25,30,35,40,45,50 的学生进行作业检查，这种抽样方法是_____.

解析：因为抽取学号是以 5 为公差的等差数列，故采用的抽样方法应是系统抽样.

答案：系统抽样



(3)某公司共有 1 000 名员工，下设若干部门，现采用分层抽样方法，从全体员工中抽取一个样本容量为 80 的样本，已告知广告部门被抽取了 4 个员工，则广告部门的员工人数为_____。

解析： $\frac{1\ 000}{80} = \frac{x}{4}$, $x=50$.

答案：50



(4)某学校高一、高二、高三年级的学生人数之比为 3 : 3 : 4, 现用分层抽样的方法从该校高中三个年级的学生中抽取容量为 50 的样本, 则应从高二年级抽取_____名学生.

解析: 设应从高二年级抽取 x 名学生, 则 $\frac{x}{50} = \frac{3}{10}$.

解得 $x=15$.

答案: 15



[全析考法]

考点一 简单随机抽样

1. 抽签法的步骤

第一步，将总体中的 N 个个体编号；

第二步，将这 N 个号码写在形状、大小相同的号签上；

第三步，将号签放在同一不透明的箱中，并搅拌均匀；

第四步，从箱中每次抽取 1 个号签，连续抽取 k 次；

第五步，将总体中与抽取的号签的编号一致的 k 个个体取出。



2. 随机数法的步骤

第一步，将个体编号；

第二步，在随机数表中任选一个数开始；

第三步，从选定的数开始，按照一定抽样规则在随机数表中选取数字，取足满足要求的数字就得到样本的号码。

[例 1] (1)总体由编号为 **01, 02, …, 19, 20** 的 **20** 个个体组成. 利用下面的随机数表选取 **5** 个个体, 选取方法是从随机数表第 **1** 行的第 **5** 列和第 **6** 列数字开始由左到右依次选取两个数字, 则选出来的第 **5** 个个体的编号为 ()

7816	6572	0802	6314	0702	4369	9728	0198
3204	9234	4935	8200	3623	4869	6938	7481

A. 08

B. 07

C. 02

D. 01

(2)下列抽取样本的方式不属于简单随机抽样的有_____.

①从无限多个个体中抽取 **100** 个个体作为样本.

②盒子里共有 **80** 个零件,从中选出 **5** 个零件进行质量检验.在抽样操作时,从中任意拿出一个零件进行质量检验后再把它放回盒子里.

③从 **20** 件玩具中一次性抽取 **3** 件进行质量检验.

④某班有 **56** 名同学,指定个子最高的 **5** 名同学参加学校组织的篮球赛.



[解析] (1)由题意知前 5 个个体的编号为 08,02,14,07,01.

(2)①不是简单随机抽样. 因为不满足总体的有限性.

②不是简单随机抽样. 因为它是放回抽样.

③不是简单随机抽样. 因为这是“一次性”抽取, 而不是“逐个”抽取.

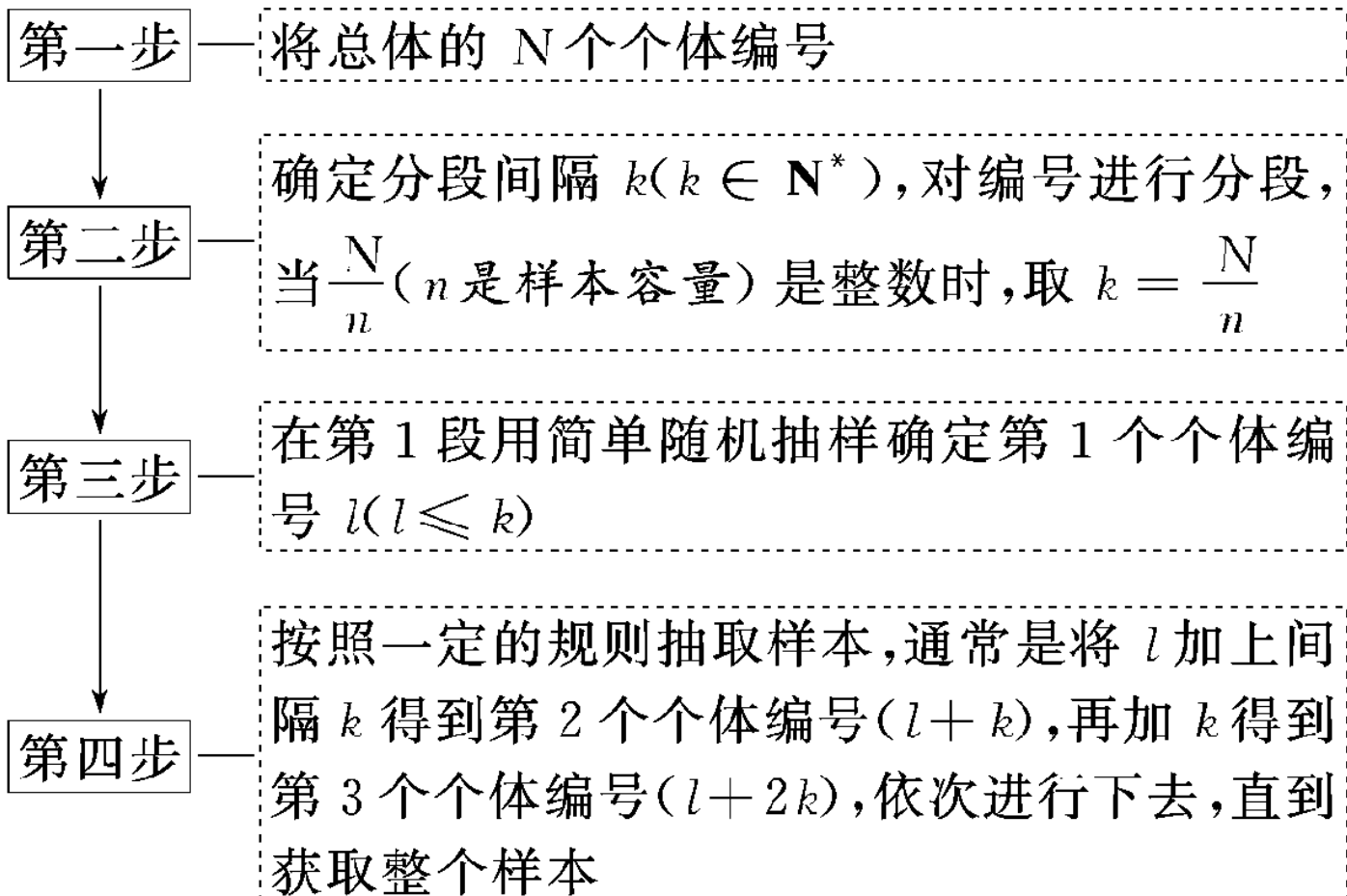
④不是简单随机抽样. 因为指定个子最高的 5 名同学是 56 名中特指的, 不存在随机性, 不是等可能抽样.

[答案] (1)D (2)①②③④



考点二 系统抽样

系统抽样的步骤



[例 2] (1)为了解 1 000 名学生的学习情况,采用系统抽样的方法,从中抽取容量为 40 的样本,则分段的间隔为 ()

A. 50

B. 40

C. 25

D. 20

(2)将高一(九)班参加社会实践编号为 1,2,3, …, 48 的 48 名学生,采用系统抽样的方法抽取一个容量为 4 的样本,已知 5 号, 29 号, 41 号学生在样本中,则样本中还有一名学生的编号是_____.



[解析] (1)由系统抽样的定义知,分段间隔为 $\frac{1\ 000}{40}=25$.

故选 C.

(2)根据系统抽样的概念,所抽取的 4 个样本的编号应成等差数列,因为在这组数中的间距为 $41-29=12$,所以所求的编号为 $5+12=17$.

[答案] (1)C (2)17



[易错提醒]

用系统抽样法抽取样本，当 $\frac{N}{n}$ 不为整数时，取 $k = \left[\frac{N}{n} \right]$ ，

即先从总体中用简单随机抽样的方法剔除 $(N - nk)$ 个个体，且剔除多余的个体不影响抽样的公平性。



进行分层抽样的相关计算时，常利用以下关系式巧解：

$$(1) \frac{\text{样本容量 } n}{\text{总体的个数 } N} = \frac{\text{该层抽取的个体数}}{\text{该层的个体数}};$$

(2) 总体中某两层的个体数之比 = 样本中这两层抽取的个体数之比.

[例 3] (1)(2018·南昌模拟)某校为了解学生学习的情况，采用分层抽样的方法从高一 1 000 人、高二 1 200 人、高三 n 人中，抽取 81 人进行问卷调查. 已知高二被抽取的人数为 30，那么 $n =$ ()

A. 860

B. 720

C. 1 020

D. 1 040



(2)(2017·江苏高考)某工厂生产甲、乙、丙、丁四种不同型号的产品，产量分别为 **200,400,300,100** 件. 为检验产品的质量，现用分层抽样的方法从以上所有的产品中抽取 **60** 件进行检验，则应从丙种型号的产品中抽取_____件.

(3)某学校三个兴趣小组的学生人数分布如下表(每名同学只参加一个小组)(单位：人).

	篮球组	书画组	乐器组
高一	45	30	a
高二	15	10	20

学校要对这三个小组的活动效果进行抽样调查，按小组分层抽样的方法，从参加这三个兴趣小组的学生中抽取 **30** 人，结果篮球组被抽出 **12** 人，则 a 的值为_____.



[解析] (1)根据分层抽样方法, 得 $\frac{1\ 200}{1\ 000+1\ 200+n} \times 81$

$=30$, 解得 $n=1\ 040$. 故选 D.

(2)本题考查分层抽样方法及用样本估计总体.

从丙种型号的产品中抽取的件数为

$$60 \times \frac{300}{200+400+300+100} = 18.$$

(3)由题意知 $\frac{12}{45+15} = \frac{30}{45+15+30+10+a+20}$, 解得 a

$=30$.

[答案] (1)D (2)18 (3)30



[方法技巧]

分层抽样的解题策略

(1) 分层抽样中分多少层, 如何分层要视具体情况而定, 总的原则是: 层内样本的差异要小, 两层之间的样本差异要大, 且互不重叠.

(2) 为了保证每个个体等可能入样, 所有层中每个个体被抽到的可能性相同.

(3) 在每层抽样时, 应采用简单随机抽样或系统抽样的方法进行抽样.

$$(4) \text{抽样比} = \frac{\text{样本容量}}{\text{总体容量}} = \frac{\text{各层样本数量}}{\text{各层个体数量}}$$



[全练题点]

1.[考点一]某工厂的质检人员对生产的 100 件产品，采用随机数法抽取 10 件检查，对 100 件产品采用下面的编号方法：

- ①1,2,3, ..., 100; ②001,002, ..., 100;
③00,01,02, ..., 99; ④01,02,03, ..., 100.

其中正确的序号是 ()

- A. ②③④ B. ③④
C. ②③ D. ①②

解析：根据随机数法编号可知，①④编号位数不统一。

答案：C



2.[考点一、二、三]对一个容量为 N 的总体抽取容量为 n 的样本，当选取简单随机抽样、系统抽样和分层抽样三种不同方法抽取样本时，总体中每个个体被抽中的概率分别为 p_1, p_2, p_3 ，则()

A. $p_1 = p_2 < p_3$

B. $p_2 = p_3 < p_1$

C. $p_1 = p_3 < p_2$

D. $p_1 = p_2 = p_3$

解析：由于三种抽样过程中，每个个体被抽到的概率都是相等的，因此 $p_1 = p_2 = p_3$.

答案： D



3.[考点二]某班共有 52 人，现根据学生的学号，用系统抽样的方法，抽取一个容量为 4 的样本，已知 3 号、29 号、42 号学生在样本中，那么样本中还有一个学生的学号是()

A. 10

B. 11

C. 12

D. 16

解析：从被抽中的 3 名学生的学号中可以看出学号间距为 13，所以样本中还有一个学生的学号是 16，故选 D.

答案： D



4.[考点三]某校高一年级有学生 400 人，高二年级有学生 360 人，现采用分层抽样的方法从全校学生中抽取 55 人，其中从高一年级学生中抽取 20 人，则从高三年级学生中抽取的人数为_____.

解析： 设从高二年级学生中抽取 x 人，由题意得 $\frac{x}{360} = \frac{20}{400}$ ，解得 $x = 18$ ，则从高三年级学生中抽取的人数为 $55 - 20 - 18 = 17$ 人.

答案： 17



5.[考点二]为了了解本班学生对网络游戏的态度，高三(6)班计划在全班 60 人中展开调查，根据调查结果，班主任计划采用系统抽样的方法抽取若干名学生进行座谈，为此先对 60 名学生进行编号为：01,02,03, …, 60，已知抽取的学生中最小的两个编号为 03,09，则抽取的学生中最大的编号为_____.

解析：由最小的两个编号为 03,09 可知，抽取时的分段间隔是 6.即抽取 10 名同学，其编号构成首项为 3，公差为 6 的等差数列，故最大编号为 $3+9\times 6=57$.

答案：57



突破点(二) 用样本预计总体

02



[基本知识]

1. 频率分布直方图和茎叶图

(1) 作频率分布直方图的步骤

①求极差(即一组数据中最大值与最小值的差); ②决定组距与组数; ③将数据分组; ④列频率分布表; ⑤画频率分布直方图.



(2) 频率分布折线图和总体密度曲线

频率分布折线图	<ul style="list-style-type: none">• 连接频率分布直方图中各小长方形上端_____，就得到频率分布折线图
总体密度曲线	<ul style="list-style-type: none">• 伴随样本容量增加，作图时所分组数_____增加，组距_____减小，对应频率折线图会越来越靠近于一条光滑曲线，统计中称这条光滑曲线为总体密度曲线

(3) 茎叶图的优点

茎叶图的优点是可以保留原始数据，而且可以随时记录，这对数据的记录和表示都能带来方便。



2. 样本的数字特征

(1) 众数、中位数、平均数

数字特征	定义与求法	优点与缺点
众数	一组数据中重复出现次数 <u>最多</u> 的数	众数体现了样本数据的最大集中点，不受极端值的影响。但显然它对其他数据信息的忽视使得无法客观地反映总体特征
中位数	把一组数据按 <u>从小到大</u> <u>大的</u> 顺序排列，处在 <u>中间</u> 位置的一个数据 (或两个数据的平均数)	中位数等分样本数据所占频率，它不受少数几个极端值的影响，这在某些情况下是优点，但它对极端值的不敏感有时也会成为缺点
平均数	如果有 n 个数据 x_1, x_2, \dots, x_n ，那么这 <u>n</u> 个数的平均数 $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$	平均数与每一个样本数据有关，可以反映出更多的关于样本数据全体的信息，但平均数受数据中的极端值的影响较大，使平均数在估计总体时可靠性降低



(2)标准差、方差

①标准差：样本数据到平均数的一种平均距离，一般用 s

表示， $s = \sqrt{\frac{1}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2]}$

②方差：标准差的平方 $s^2 = \frac{1}{n}[(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2]$ ，其中 $x_i (i=1,2,3, \dots, n)$ 是样本数据， n 是样本容量， \bar{x} 是样本平均数。

③方差与标准差相比，都是衡量样本数据离散程度的统计量，但方差因为对标准差进行了平方运算，夸大了样本的偏差程度。

(3)平均数、方差公式的推广

若数据 x_1, x_2, \dots, x_n 的平均数为 \bar{x} ，方差为 s^2 ，则数据 $mx_1 + a, mx_2 + a, \dots, mx_n + a$ 的平均数为 $m\bar{x} + a$ ，方差为 m^2s^2 。

[基本能力]

1. 判断题

(1)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.

(✓)

(2)在频率分布直方图中，众数左边和右边的小长方形的面积和是相等的.

(✗)

(3)从频率分布直方图得不出原始的数据内容，把数据表示成直方图后，原有的具体数据信息就被抹掉了.

(✓)

(4)茎叶图一般左侧的叶按从大到小的顺序写，右侧的叶按从小到大的顺序写，相同的数据可以只记一次.

(✗)

(5)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.

(✓)

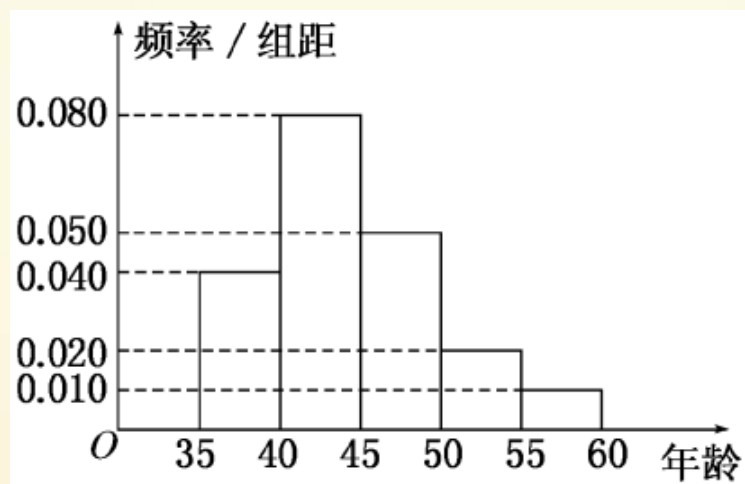
(6)一组数据的众数可以是一个或几个，中位数也具有相同的结论.

(✗)



2. 填空题

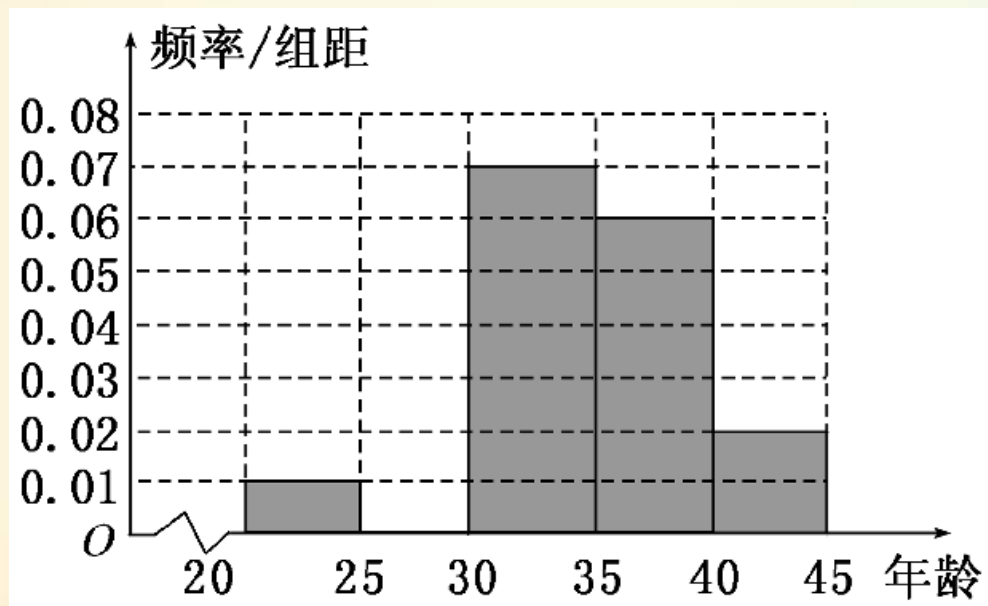
(1)某校为了了解教科研工作开展状况与教师年龄之间的关系，将该校不小于35岁的80名教师按年龄分组，分组区间为[35,40)，[40,45)，[45,50)，[50,55)，[55,60]，由此得到频率分布直方图如图，则这80名教师中年龄小于45岁的有_____人。



解析：由频率分布直方图可知45岁以下的教师的频率为 $5 \times (0.040 + 0.080) = 0.6$ ，所以共有 $80 \times 0.6 = 48$ (人)。

答案：48

(2)对某市“四城同创”活动中 800 名志愿者的年龄抽样调查统计后得到频率分布直方图(如图), 但是年龄组为 $[25,30)$ 的数据不慎丢失, 则依据此图可得:



① $[25,30)$ 年龄组对应小矩形的高度为_____;

②据此估计该市“四城同创”活动中志愿者年龄在 $[25,35)$ 的人数为_____.



解析： 设 $[25,30)$ 年龄组对应小矩形的高度为 h ，则 $5 \times (0.01 + h + 0.07 + 0.06 + 0.02) = 1$ ，解得 $h = 0.04$. 则志愿者年龄在 $[25,35)$ 年龄组的频率为 $5(0.04 + 0.07) = 0.55$ ，故志愿者年龄在 $[25,35)$ 年龄组的人数约为 $0.55 \times 800 = 440$.

答案： ①0.04 ②440



(3)对某商店一个月内每天的顾客人数进行了统计,得到样本的茎叶图(如图所示),则该样本的中位数、众数、极差分别是_____.

1	2 5
2	0 2 3 3
3	1 2 4 4 8 9
4	5 5 5 7 7 8 8 9
5	0 0 1 1 4 7 9
6	1 7 8

解析: 由题意知各数为 12,15,20,22,23,23,31,32,34,34,38,39,45,45,45,47,47,48,48,49,50,50,51,51,54,57,59,61,67,68, 中位数是 46, 众数是 45, 最大数为 68, 最小数为 12, 极差为 $68-12=56$.

答案: 46,45,56



(4)一组数据分别为：12,16,20,23,20,15,28,23，则这组数据的中位数是_____.

解析：这组数据从小到大排列为：12,15,16,20,20,23,23,28，

∴这组数据的中位数是 $\frac{20+20}{2}=20$.

答案：20



(5) 已知一组数据 4.7, 4.8, 5.1, 5.4, 5.5, 则该组数据的方差是 _____.

解析: 5 个数的平均数 $\bar{x} = \frac{4.7 + 4.8 + 5.1 + 5.4 + 5.5}{5} = 5.1$, 所

以它们的方差 $s^2 = \frac{1}{5}[(4.7 - 5.1)^2 + (4.8 - 5.1)^2 + (5.1 - 5.1)^2 + (5.4 - 5.1)^2 + (5.5 - 5.1)^2] = 0.1$.

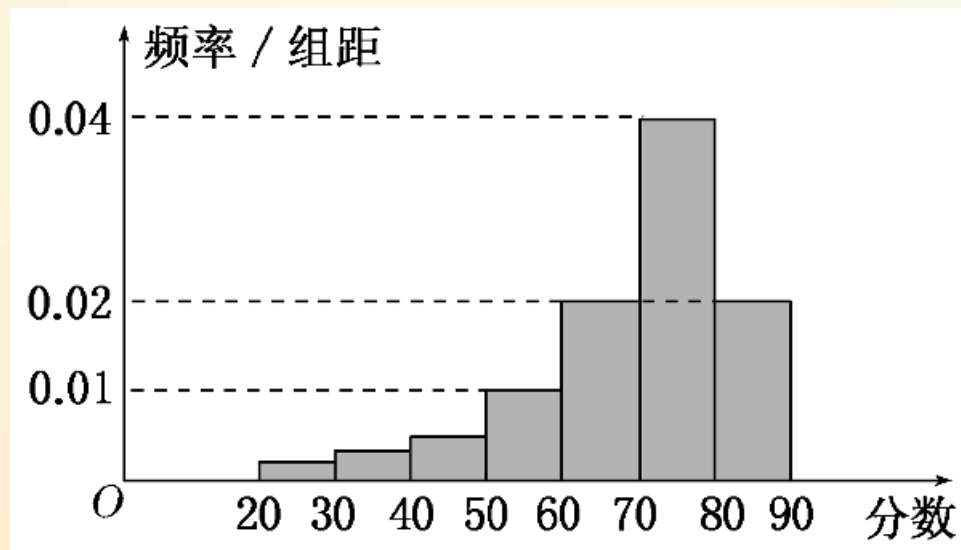
答案: 0.1



[全析考法]

考点一 频率分布直方图

[例 1] (2017·北京高考)某大学艺术专业 400 名学生参加某次测评, 根据男女学生人数比例, 使用分层抽样的方法从中随机抽取了 100 名学生, 记录他们的分数, 将数据分成 7 组: $[20,30)$, $[30,40)$, ..., $[80,90]$, 并整理得到如下频率分布直方图:



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/207122041102006063>