

# 统计

## 一、统计学中的基本概念

### 1. 知识讲解

#### 总体

统计中所考察对象的某一数值指标的全体构成的集合称为总体；

#### 个体

构成总体的每一个元素叫做个体；

#### 样本

从总体中抽取若干个个体进行考察，这若干个个体所构成的集合叫做总体的一个样本。

#### 样本容量

样本中包含个体的数量称为样本容量。

### 2. 例题精讲

1 为调查参加运动会的1000名运动员的年龄情况，从中抽查了100名运动员的年龄，就这个问题来说，下列说法正确的是（ ）。

A. 1000名运动员是总体

B. 每个运动员是个体

C. 抽取的100名运动员是样本

D. 样本容量是100

### 3. 课堂巩固

2 为了了解某地参加计算机水平测试的5000名学生的成绩，从中抽取了200名学生进行调查分析．在这个问题中，被抽取的200名学生是（ ）．

- A. 总体                      B. 个体                      C. 样本                      D. 样本量

## 二、随机抽样

### 1. 知识讲解



#### 简单随机抽样

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**．



#### 分类

最常用的简单随机抽样方法有两种——**抽签法**和**随机数法**．

#### （1）抽签法（抓阄法）

##### ①定义：

抽签法就是把总体中的 $N$ 个个体编号，把号码写在号签上，将号签放在一个容器中，搅拌均匀后，每次从中抽取一个号签，连续抽取 $n$ 次，就得到一个容量为 $n$ 的样本．

##### ②步骤：

a. 编号：即给总体中的所有个体编号，号码可以从1到 $N$ ；

b. 制签：即将1~ $N$ 这 $N$ 个号码写在形状、大小相同的号签上（号签可以用小球、卡片、纸条等制作）；

c. 搅拌：即将号签放在一个不透明的容器中，充分搅拌至均匀；

d. 抽取：即从容器中每次抽取一个号签（不放回抽取），并记录其编号，连续抽取 $n$ 次．

## (2) 随机数表法

### ①定义：

随机数表是由0, 1, 2, ..., 9这10个数字组成的数表，并且表中的每一位置出现各个数字的可能性相同。通过随机数表，根据实际需要和方便使用的原则，将几个数组合成一组，然后通过随机数表抽取样本。

### ②步骤：

- a. 将总体中的所有个体进行编号（每个号码位数一致）；
- b. 在随机数表中任选一个数作为起始号码；
- c. 从选定的数开始按照一定的方向读下去，获取与事先编号位数一致的号码；若得到的号码不在总体内，则跳过；若在总体中，则取出；如果得到的号码前面已经出现过，也跳过，如此继续下去，直到取得满足样本容量为止；

## 2. 例题精讲

3 抽签法中确保样本代表性的关键是（ ）。

- A. 制签                      B. 搅拌均匀                      C. 逐一抽取                      D. 抽取不放回

4 从某单位50名职工中随机抽取5名职工参加一项社区服务活动，用随机数法确定这5名职工，现将随机数表摘录部分如下：

|    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 16 | 22 | 77 | 94 | 39 | 49 | 54 | 43 | 54 | 81 | 77 | 37 | 93 | 23 | 78 | 87 | 35 | 20 | 96 | 43 |
| 84 | 42 | 17 | 53 | 31 | 57 | 24 | 55 | 06 | 88 | 77 | 04 | 74 | 47 | 67 | 21 | 76 | 33 | 50 | 25 |

从随机数表第一行的第6列和第7列数字开始由左到右依次选取两个数字，则选出的第5个职工的编号为（ ）。

- A. 23                      B. 37                      C. 35                      D. 17

## 3. 课堂巩固

5 关于简单随机抽样，下列说法正确的是（ ）

- ①它要求被抽取样本的总体的个数有限；

②它是从总体中逐个地进行抽取；

③它是一种不放回抽样；

④它是一种等可能性抽样。

A. ①②③④

B. ③④

C. ①②③

D. ①③④

6 某工厂利用随机数表对生产的600个零件进行抽样测试，先将600个零件进行编号，编号分别为001, 002, ..., 599, 600从中抽取60个样本，如下提供随机数表的第4行到第6行：

32 21 18 34 29 78 64 54 07 32 52 42 06 44 38 12 23 43 56 77 35 78 90 56 42  
84 42 12 53 31 34 57 86 07 36 25 30 07 32 86 23 45 78 89 07 23 68 96 08 04  
32 56 78 08 43 67 89 53 55 77 34 89 94 83 75 22 53 55 78 32 45 77 89 23 45

若从表中第6行第6列开始向右依次读取3个数据，则得到的第6个样本编号（ ）。

A. 522

B. 324

C. 535

D. 578



### 系统抽样

当总体中的个体数较多时，可将总体分成均衡的几个部分，然后按照预先定出的规则，从每一部分抽取1个个体，得到所需要的样本，这种抽样方法叫做**系统抽样**。



### 步骤

假设现在要从容量为 $N$ 的总体中抽取容量为 $n$ 的样本，我们可以按下列步骤进行系统抽样：

(1) 先将总体的 $N$ 个个体编号。有时可直接利用个体自身所带的号码，如学号、准考证号、门牌号等；

(2) 确定分段间隔，对编号进行分段。当 $\frac{N}{n}$  ( $n$ 是样本容量)是整数时，取 $k = \frac{N}{n}$ ；

(3) 在第一段用简单随机抽样确定第一个个体编号 $l$  ( $l \leq k$ )；

(4) 按照一定的规则抽取样本。通常是将 $l$ 加上间隔 $k$ 得到第二个个体编号 $(l + k)$ ，再加上 $k$ 得到第三个个体编号 $(l + 2k)$ ，依次进行下去，直到获取整个样本。

#### 【补充说明】

(1) 按照系统抽样的要求，按照先后顺序抽取得到的号码成等差数列，公差为常数，这个常数就是分段间隔，因此系统抽样又称为等距抽样；

(2) 如果遇到  $\frac{N}{n}$  不是整数的情况, 可以先从总体中随机地剔除几个个体, 使得总体中剩余的个体数能被样本容量整除.

#### 4. 例题精讲

7 若总体中含有1655个个体, 现在要用系统抽样法从中抽取一个容量为35的样本, 分段时应从总体中随机剔除 \_\_\_\_\_ 个个体, 编号后应均分为 \_\_\_\_\_ 段, 每段由 \_\_\_\_\_ 个个体.

8 某班有学生的60人, 将这60名学生随机编号为1~60号, 用系统抽样的方法从中抽出4名学生, 已知3号、33号、48号学生在样本中, 则样本中另一个学生的编号为 ( ).

A. 28

B. 23

C. 18

D. 13

9 要从编号为1~50的50枚最新研制的某种型号的导弹中抽取5枚进行发射试验, 用系统抽样方法确定所选取的5枚导弹的编号, 下面所给样本中, 只能是 ( ).

A. 5, 10, 15, 20, 25

B. 3, 13, 23, 33, 43

C. 1, 2, 3, 4, 5

D. 2, 4, 8, 16, 32

10 最近网络上流行一种“农场游戏”, 这种游戏通过虚拟软件模拟种植与收获的过程. 为了解本班学生对此游戏的态度, 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, 对此先对60名学生进行编号, 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是 \_\_\_\_\_.

#### 5. 课堂巩固

11 高二某班共有学生56人, 座号分别为1, 2, 3, ..., 56, 现根据座号, 用系统抽样的方法, 抽取一个容量为4的样本. 已知5号、19号、47号同学在样本中, 那么样本中还有一个同学的座号是 \_\_\_\_\_.

12 有60件产品，编号为1至60，现从中抽取5件进行检验，用系统抽样的方法所确定的抽样编号是（ ）。

A. 5, 10, 15, 20, 25

B. 5, 12, 31, 39, 57

C. 5, 15, 25, 35, 45

D. 5, 17, 29, 41, 53

13 从编号为001, 002, ..., 500的500个产品中用系统抽样的方法抽取一个样本，已知样本中最小的两个编号分别为007, 032，则样本容量为（ ）。

A. 10

B. 15

C. 20

D. 25

14 从编号为0, 1, 2, ..., 49的50件产品中，采用系统抽样的方法抽取容量是5份样本，若编号为27的产品在样本中，则该样本中产品的最大编号为 \_\_\_\_\_。



### 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，这种抽样方法叫做**分层抽样**。



### 步骤

- ①将总体按一定标准进行分层；
- ②计算各层的个体数与总体的个体数的比值；
- ③按各层个体数占总体个体数的比值确定各层应抽取的容量；
- ④在每一层进行抽样（采取简单随机抽样或系统抽样）。

## 6. 例题精讲

15 某校共有高一、高二、高三学生1290人，其中高一480人，高二比高三多30人，为了解该校学生的身体健康情况，现采用分层抽样方法进行调查，在抽取的样本中有高一学生96人，则该样本中的

已知某高中共有2400人，其中高一年级600人，现对该高中全体学生利用分层抽样的方法进行一项调查，需要从高一年级抽取30人，则全校应一共抽取 \_\_\_\_\_ 人．

17 某初级中学有学生270人，其中一年级108人，二、三年级各81人，现要利用抽样方法取10人参加某项调查，考虑选用简单随机抽样、分层抽样和系统抽样三种方案，使用简单随机抽样和分层抽样时，将学生按一、二、三年级依次统一编号为1, 2, ..., 270；使用系统抽样时，将学生统一随机编号1, 2, ..., 270，并将整个编号依次分为10段；如果抽得号码有下列四种情况：

- ① 7, 34, 61, 88, 115, 142, 169, 196, 223, 250 ;
- ② 5, 9, 100, 107, 111, 121, 180, 195, 200, 265 ;
- ③ 11, 38, 65, 92, 119, 146, 173, 200, 227, 254 ;
- ④ 30, 57, 84, 111, 138, 165, 192, 219, 246, 270 ;

关于上述样本的下列结论中，正确的是（ ）。

- A. ②、③都不能为系统抽样  
B. ②、④都不能为分层抽样  
C. ①、④都可能为系统抽样  
D. ①、③都可能为分层抽样

18 现要完成下列3项抽样调查：

- ①从15种疫苗中抽取5种检测是否合格．
- ②某中学共有480名教职工，其中一线教师360名，行政人员48名，后勤人员72名．为了解教职工对学校校务公开方面的意见，拟抽取一个容量为20的样本．
- ③某中学报告厅有28排，每排有35个座位，一次报告会恰好坐满了听众，报告会结束后，为了听取意见，需要请28名听众进行座谈．

较为合理的抽样方法是( )。

- A. ①简单随机抽样, ②分层抽样, ③系统抽样 B. ①简单随机抽样, ②系统抽样, ③分层抽样  
C. ①系统抽样, ②简单随机抽样, ③分层抽样 D. ①分层抽样, ②系统抽样, ③简单随机抽样

## 7. 课堂巩固

19 某单位有职工100人，不到35岁的有45人，35岁到49岁的有25人，剩下的为50岁以上(包括50岁)的人，用分层随机抽样的方法从中抽取20人，各年龄段分别抽取的人数为( )。

A. 7, 5, 8

B. 9, 5, 6

C. 6, 5, 9

D. 8, 5, 7

20 某校高一年级有 $x$ 个学生，高二年级有 $y$ 个学生，高三年级有 $z$ 个学生。采用分层抽样的方法抽取一个容量为45的样本，高一年级被抽取20人，高二年级被抽取10人，若高三年级共有300人，则此学校共有 \_\_\_\_\_ 人。

21 某公司在甲、乙、丙、丁四个地区分别有150个、120个、180个、150个销售点。公司为了调查产品销售部的情况，需从这600个销售点中抽取一个容量为100的样本，记这项调查为①；

在丙地区中有20个特大型销售点，要从中抽取7个调查其销售收入和售后服务情况，记这项调查为②；

一个年级有10个班，每个班有50名同学，随机编为1至50号，为了了解他们的学习情况，要求每个班的30号同学留下来进行问卷调查，记这项调查为③。

则完成①、②、③这三项调查宜采用的抽样方法依次是( )。

A. 分层抽样法，系统抽样法，简单随机抽样

B. 分层抽样法，简单随机抽样法，系统抽样

C. 系统抽样法，分层抽样法，简单随机抽样

D. 简单随机抽样法，分层抽样法，系统抽样

### 三、图表法估计总体分布

#### 1. 知识讲解



#### 频数与频率

将一批数据额按要求分成若干组，各组内数据的个数，叫做该组的频数。每组频数除以全体数据个数所得到的结果叫做该组的频率，由频率的计算方法可知，频率表示该组数据在样本中所占比例的大小。



#### 频率分布

由各个小组数据在样本容量中所占比例的大小反映出这些数据的分布规律，称为样本数据的**频率分布**。

## 频率分布表

### (1) 概念

为了能直观地显示样本的频率分布情况，通常将样本的容量、样本中出现该事件的频数以及计算所得的相应频率列在一张表中，这张表就叫做样本的频率分布表。

### (2) 频率分布表的绘制步骤

①求极差（极差通常以大写字母  $R$  表示）：

极差即为一组数据中最大值与最小值的差，它反映的数据的上下限和波动范围；

②决定组距与组数：

组距（ $c$ ）是指每个小组的两个端点之间的距离。通常来说，组距与组数的确定并没有固定的标准。我们可以通过下面的原则来合理选择组距与组数：

若  $\frac{N}{c}$  是整数，则组数 =  $\frac{N}{c}$ ；

若  $\frac{N}{c}$  不是整数，则组数 =  $\left[ \frac{N}{c} \right] + 1$ （ $[x]$  表示不超过  $x$  的最大整数）。

③将数据分组

通常对组内数值所在区间取左闭右开区间，最后一组取闭区间；

④列频率分布表

计算各小组的频率，将相应内容填入下表：

| 分组                   | 频数        | 频率        |
|----------------------|-----------|-----------|
| $[a_0, a_1)$         | $N_1$     | $f_1$     |
| $[a_1, a_2)$         | $N_2$     | $f_2$     |
| $[a_2, a_3)$         | $N_3$     | $f_3$     |
| .....                | .....     | .....     |
| $[a_{n-2}, a_{n-1})$ | $N_{n-1}$ | $f_{n-1}$ |
| $[a_{n-1}, a_n]$     | $N_n$     | $f_n$     |
| 合计                   | $N$       | 1         |

【补充说明】

(1) 可以利用 
$$\begin{cases} N = \sum_{i=1}^n N_i \\ 1 = \sum_{i=1}^n f_i \end{cases}$$
 来检验制表是否正确；

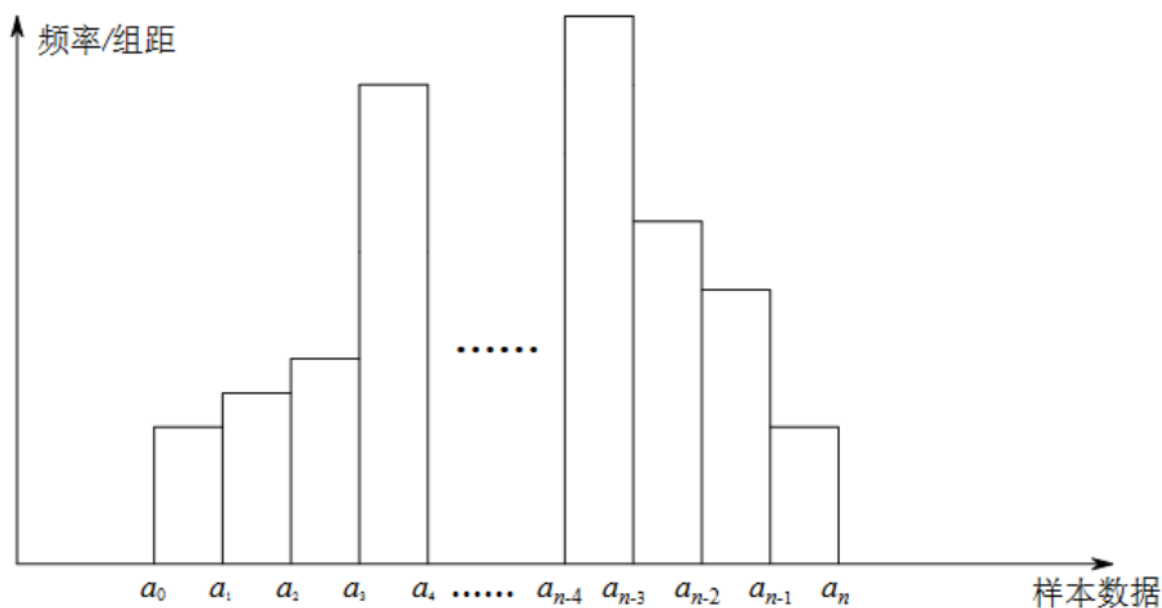
(2) 当样本容量不超过100时，按照数据的多少，通常分成5 ~ 12组。

### 频率分布直方图

将频率分布表中的结果以柱状图的形式直观形象的表现出来的图形称为频率分布直方图。

### 频率分布直方图的绘制步骤

- (1) 根据频率分布表确定组距 ( $c$ ) 与组数，将组距作为小矩形的底；
- (2) 根据频率 ( $f$ ) 与组距 ( $c$ ) 确定矩形的高，高 =  $\frac{f}{c}$ ；
- (3) 依据频率分布表以及确定的矩形的大小绘制频率分布直方图，如下图：

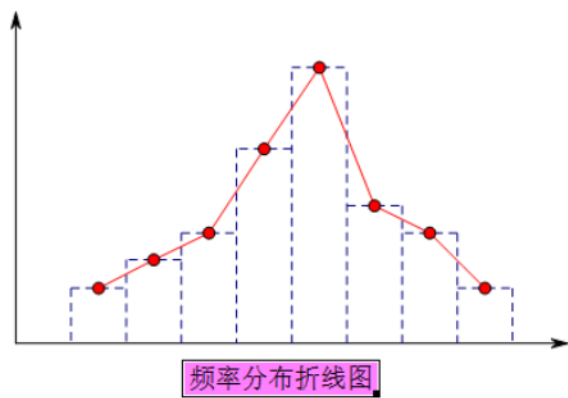
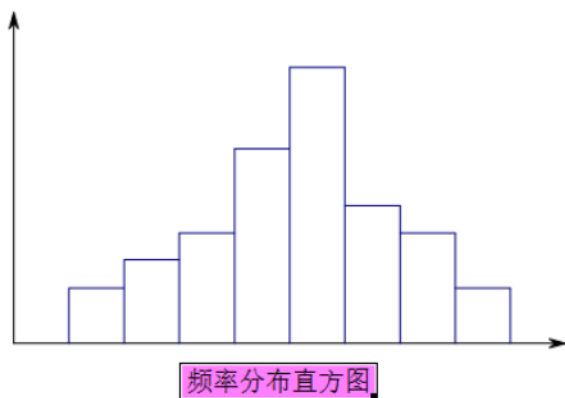


#### 【补充说明】

- (1) 各个小长方形的面积表示相应各组的频率。这样，频率分布直方图就以面积的形式反映了数据落在各个小组的频率的大小；
- (2) 容易知道，在频率分布直方图中，各小长方形的面积的总和为1。

### 频率分布折线图

连接频率分布直方图中各小长方形上端的中点，就得到频率分布折线图，如下图：



## 茎叶图

除了上面几种图表能帮助我们理解样本数据外，统计学中还有一种被用来表示数据的图叫做茎叶图。

我们结合下面的例子来说明制作茎叶图的方法，以及从茎叶图中提取样本数据信息的方法。

某赛季甲、乙两名篮球运动员每场比赛得分的原始记录如下：

甲运动员得分：13, 51, 23, 8, 26, 38, 16, 33, 14, 28, 39；

乙运动员得分：49, 24, 12, 31, 50, 31, 44, 36, 15, 37, 25, 36, 39。

用茎叶图表示如下：

| 甲 |   |   |   | 乙 |   |         |
|---|---|---|---|---|---|---------|
|   | 8 |   | 0 |   |   |         |
| 4 | 6 | 3 | 1 | 2 | 5 |         |
| 3 | 6 | 8 | 2 | 5 | 4 |         |
| 3 | 8 | 9 | 3 | 1 | 6 | 1 6 7 9 |
|   |   |   | 4 | 4 | 9 |         |
|   | 1 |   | 5 | 0 |   |         |

顾名思义，茎是指中间的一列数，叶就是从茎的旁边生长出来的数。中间的数字表示得分的十位数，旁边的数字分别表示两个人得分的个位数。

从上图可以看出，茎叶图不仅能够保留原始数据，而且能够展示数据的分布情况。比如，乙运动员的得分基本上是对称的，叶的分布是“单峰”的，有 $\frac{10}{13}$ 的叶集中在茎2, 3, 4上，中位数是36；甲运动员的得分除一个特殊得分（51分）外，也大致对称，叶的分布也是“单峰”的，有 $\frac{9}{11}$ 的叶主要集中在茎1, 2, 3上，中位数是26。由此可以看出，乙运动员的成绩更好。另外，从叶在茎的分布情况

看，以运动员的得分更集中于峰值附近，这说明乙运动员的发挥更稳定．

### 【补充说明】

(1) 需要注意的是茎叶图的“叶”，无论数据位数是几位，每个数据只能保留一个数字．因此当数据的位数多于两位时，“茎”的数据不再是一位．例如对于下组数据：

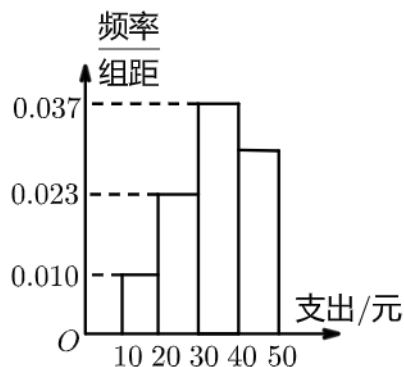
10.7, 10.5, 11.3, 11.6, 13.4, 16.6, 16.7, 17.0, 17.3

茎的位置对应的数据应该是10, 11, 13, 16, 17．

(2) 表示茎的数字按照由上到下递增的顺序排列；对于两位数来说，十位数字充当“茎”，各位数字充当“叶”．对于小数来说，整数部分充当“茎”，小数部分充当“叶”．解题时要根据数据的特点合理选择“茎”和“叶”．

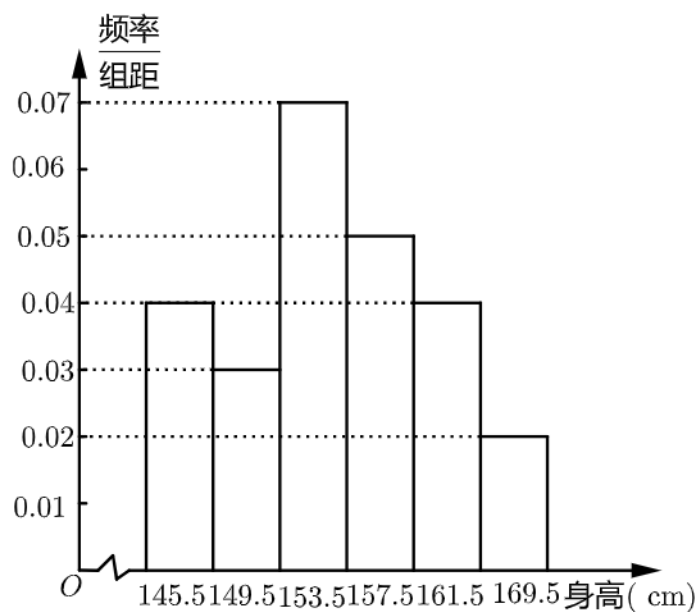
## 2. 例题精讲

- 22 学校为了解学生每月在课外读物方面的支出情况，抽取了 $n$ 个同学进行调查，结果显示这些同学的支出（单位：元）都在 $[10, 50]$ 内，其频率分布直方图如图所示，其中支出在 $[30, 50]$ 内的同学有67人，则 $n = \underline{\hspace{2cm}}$ ．



- 23 某中学高一女生共有450人，为了了解高一女生的身高（单位：cm）情况，随机抽取部分高一女生测量身高，所得数据整理后列出频率分布表如下：

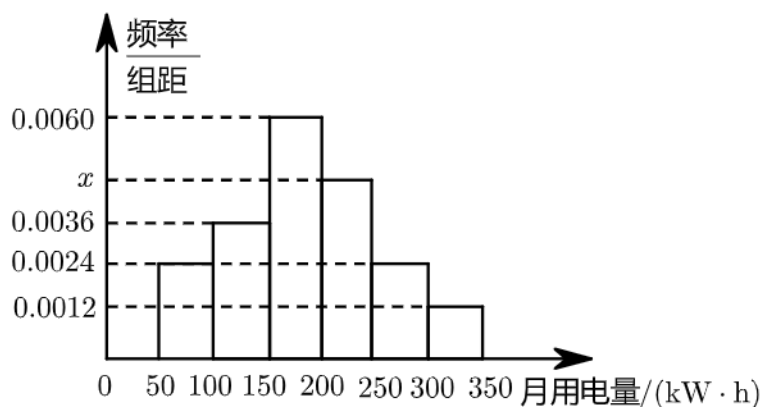
| 组别               | 频数  | 频率   |
|------------------|-----|------|
| $[145.5, 149.5)$ | 8   | 0.16 |
| $[149.5, 153.5)$ | 6   | 0.12 |
| $[153.5, 157.5)$ | 14  | 0.28 |
| $[157.5, 161.5)$ | 10  | 0.20 |
| $[161.5, 165.5)$ | 8   | 0.16 |
| $[165.5, 169.5]$ | $m$ | $n$  |
| 合计               | $M$ | $N$  |



(1) 求出表中字母  $m = \underline{\hspace{1cm}}$  ,  $n = \underline{\hspace{1cm}}$  ,  $M = \underline{\hspace{1cm}}$  ,  $N = \underline{\hspace{1cm}}$  .

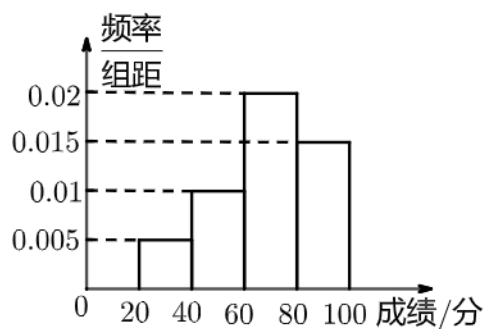
(2) 估计该校高一女生身高在  $[149.5, 165.5]$  范围内的有  $\underline{\hspace{1cm}}$  人 .

- 24 从某小区抽取100户居民用户进行月用电量调查，发现他们的用电量都在  $50 \sim 350 \text{ kW} \cdot \text{h}$  之间，进行适当分组后（每组为左闭右开的区间），画出频率分布直方图如图所示．则直方图中  $x$  的值为  $\underline{\hspace{1cm}}$  ；在被调查的用户中，用电量落在区间  $[100, 250)$  内的户数为  $\underline{\hspace{1cm}}$  .



### 3. 课堂巩固

- 25 某学校组织学生参加英语测试，成绩的频率分布直方图如图所示，数据的分组依次为 $[20, 40)$ ， $[40, 60)$ ， $[60, 80)$ ， $[80, 100]$ ，若低于60分的人数是15人，则参加英语测试的学生人数是（ ）。

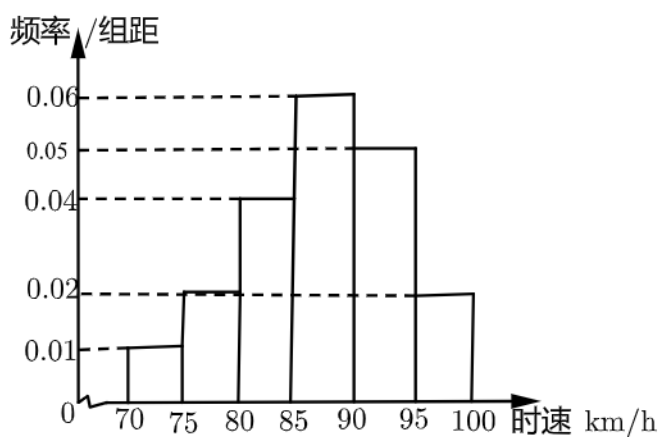


A. 45

B. 50

C. 55

- 26 港珠澳大桥于2018年10月2刻日正式通车，它是中国境内一座连接香港、珠海和澳门的桥隧工程，桥隧全长55千米。桥面为双向六车道高速公路，大桥通行限速100km/h，现对大桥某路段上1000辆汽车的行驶速度进行抽样调查。画出频率分布直方图（如图），根据直方图估计在此路段上汽车行驶速度在区间 $[85, 90)$ 的车辆数和行驶速度超过90km/h的频率分别为（ ）。



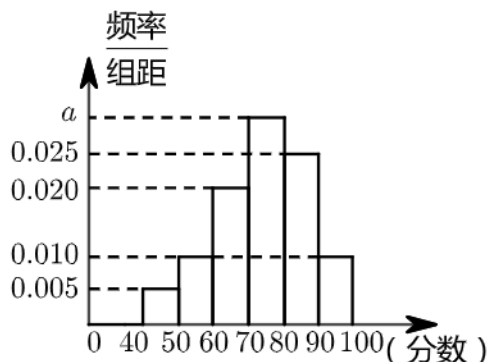
A. 300, 0.25

B. 300, 0.35

C. 60, 0.25

D. 60, 0.35

- 27 从某校高一年级学生中随机抽取了20名学生，将他们的数学检测成绩（分）分成六段（满分100分，成绩均为不低于40分的整数）： $[40, 50)$ ， $[50, 60)$ ， $\dots$ ， $[90, 100]$ 后，得到如图所示的频率分布直方图。



- (1) 求图中实数 $a$ 的值。
- (2) 若该校高一年级共有学生600名，试根据以上数据，估计该校高一年级数学检测成绩不低于80分的人数。

## 四、样本数字特征估计总体数字特征

### 1. 知识讲解



#### 百分位数

第 $p$ 百分位数（ $p\%$ 分位数）是一个数值，指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 $1 - p\%$ 的数据大于或等于这个数值。



#### 计算步骤

计算一组 $n$ 个数据的第 $p$ 百分位数：

- ①按从小到大排列原始数据；
- ②计算 $i = n \cdot p\%$ ；
- ③若 $i$ 不是整数，而大于 $i$ 的比邻整数为 $j$ ，则第 $p$ 百分位数为第 $j$ 项数据；

若 $i$ 是整数，则第 $p$ 百分位数为第 $i$ 项与第 $(i+1)$ 项数据平均值。

【补充说明】

第25百分位数、第50百分位数、第75百分位数统称为四分位数。

其中，第25百分位数又称为第一四分位数或下四分位数，第75百分位数又称为第三四分位数或上四分位数。

## 2. 例题精讲

28 下表为12名毕业生的起始月薪

| 毕业生 | 起始月薪 |
|-----|------|
| 1   | 2850 |
| 2   | 2950 |
| 3   | 3050 |
| 4   | 2880 |
| 5   | 2755 |
| 6   | 2710 |
| 7   | 2890 |
| 8   | 3130 |
| 9   | 2940 |
| 10  | 3325 |
| 11  | 2920 |
| 12  | 2880 |

根据表中所给的数据计算第85百分位数是\_\_\_\_\_。

29 对某自行车赛手在相同条件下进行了12次测试，测得其最大速度（单位： $\text{m/s}$ ）的数据如下：27，38，30，36，35，31，33，29，38，34，28，36，则他的最大速度的第一四分位数是（ ）。

A. 29

B. 29.5

C. 30

D. 36

### 3. 课堂巩固

30 高二(1)班7人宿舍中每个同学的身高分别为170, 168, 172, 172, 175, 176, 180, 求这7人的第40百分位数为( ) .

A. 168

B. 170

C. 172

D. 171

31 有一种鱼的身体吸收汞, 汞的含量超过其体重的1.00ppm (即百万分之一) 的鱼被人食用后, 就会对人产生危害, 在12条鱼的样本中发现的汞含量 (单位: ppm) 如下:

0.24   0.98   0.96   0.98   0.39   0.84

1.44   1.58   1.08   1.20   1.62   1.68

则估计这批鱼汞含量的第25, 80百分位数分别是( ) .

A. 0.96, 1.20

B. 0.84, 1.20

C. 0.90, 1.58

D. 0.84, 1.58



#### 众数

**众数**就是一组数据中出现次数最多的数值.

#### 【补充说明】

(1) 一组数据中的众数可能不止一个. 例如, 数据2, 3, 3, 4, 8, 9, 9的众数是3和9.

(2) 一组数据中如果所有子数据出现的次数都是一样的, 那么就称这组数据没有众数. 如数据1, 2, 3, 4, 5.



#### 作用

众数容易查找和计算, 但是它仅仅体现了样本数据的最大集中点, 除此之外无法反应样本其他数据信息, 作用不大.



#### 中位数

对于有限数据组, 将数据升序 (降序) 排列之后位于正中间位置的数字就是这组数据的中位数, 当数据组容量为偶数时, 取中间两个数据的平均数作为这组数据的中位数.

例如:

数据6, 9, 8, 5, 4, 3, 8, 2, 6, 4, 3升序调整为2, 3, 3, 4, 4, 5, 6, 6, 8, 8, 9, 中位数为5;

数据6, 8, 5, 4, 3, 8, 2, 6, 4, 3升序调整为2, 3, 3, 4, 4, 5, 6, 6, 8, 8, 中位数为 $\frac{4+5}{2} = 4.5$ .

#### 【补充说明】

(1) 不同于众数的是, 对于一组数据, 中位数只能有一个;

(2) 对于所有数值都一样的一组数据来说, 中位数就是这个数值. 如数据5, 5, 5, 5, 5, 5, 5, ...的中位数为5.



#### 作用

样本中位数不受少数几个极端值的影响, 这在某些情况下是一个优点, 但它对极端值的不敏感有时也会成为缺点.



#### (算数) 平均数

样本中所有数据之和与样本容量之比称为样本的平均数.

例如: 若样本为2, 3, 3, 4, 4, 5, 6, 6, 8, 8, 则平均数为 $\bar{x} = \frac{2+3+3+4+4+5+6+6+8+8}{10} = 4.9$ .



#### 计算方法

(1) 利用定义:

$$n \text{ 个数据 } x_1, x_2, x_3, \dots, x_n \text{ 的平均数 } \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

(2) 增量替代:

如果数据 $x_1, x_2, x_3, \dots, x_n$ 在某个固定常数 $a$ 附近波动, 可以先将每个数据减去常数 $a$ 得到新数据

$x_1 - a, x_2 - a, x_3 - a, \dots, x_n - a$ , 计算其平均数为 $\bar{x}'$ , 则原数据 $x_1, x_2, x_3, \dots, x_n$ 的平均数

$$\bar{x} = \bar{x}' + a.$$



#### 作用

由于样本平均数与每一个数据有关, 所以, 任何一个样本数据的改变都会引起平均数的改变. 这是众数、中位数都不具有的性质. 也正是因为这个原因, 与众数、中位数比较起来, 平均数可以反映出更多的关于样本数据全体的信息.

#### 【补充说明】

若数据  $x_1, x_2, x_3, \dots, x_n$  的平均数为  $\bar{x}$ ，则数据  $mx_1 + t, mx_2 + t, mx_3 + t, \dots, mx_n + t$  的平均数为  $m\bar{x} + t$ 。



## 标准差与方差

标准差是样本数据到平均数的一种**平均距离**，一般用  $s$  表示；标准差的平方  $s^2$  称为**方差**。



## 计算方法

所谓“平均距离”，其含义可以作如下理解：

假设样本数据是  $x_1, x_2, x_3, \dots, x_n$ ， $\bar{x}$  表示这组数据的平均数。 $x_i$  到  $\bar{x}$  的距离是

$$|x_i - \bar{x}| (i = 1, 2, \dots, n) .$$

于是，样本数据  $x_1, x_2, x_3, \dots, x_n$  到  $\bar{x}$  的“平均距离”是

$$S = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| .$$

由于上式含有绝对值，运算太不方便，因此通常改用如下公式来计算**标准差**

$$s = \sqrt{\frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} .$$

因此**方差**的计算方法为： $s^2 = \frac{1}{n} [(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2]$ 。



## 作用

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，数据的离散程度越大；标准差、方差越小，数据的离散程度越小。故借助标准差或方差可以考察一组数据的稳定情况。

### 【补充说明】

(1) 标准差和方差的取值范围为  $[0, +\infty)$ ；

(2) 方差的性质：

① 数据  $x_1, x_2, x_3, \dots, x_n$  与数据  $x_1 + a, x_2 + a, x_3 + a, \dots, x_n + a$  的方差相等；

② 若数据  $x_1, x_2, x_3, \dots, x_n$  的方差为  $s^2$ ，则数据  $ax_1, ax_2, ax_3, \dots, ax_n$  的方差为  $a^2 s^2$ 。

(3) 方差的简化计算公式：

$$s^2 = \frac{1}{n} (x_1^2 + x_2^2 + \dots + x_n^2) - (\bar{x})^2 = \overline{x^2} - (\bar{x})^2 .$$

## 4. 例题精讲

- 32 某校高二(4)班统计全班同学中午在食堂用餐时间,有7人用时为6分钟,有14人用时7分钟,有15人用时为8分钟,还有4人用时为10分钟,则高二(4)班全体同学用餐平均用时为( )分钟.

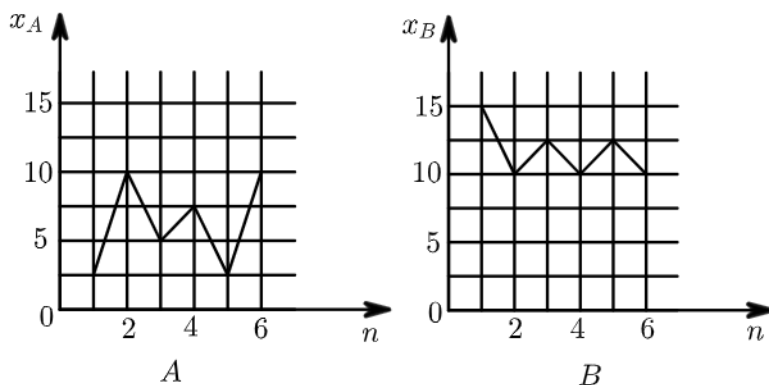
A. 7                      B. 7.5                      C. 8                      D. 8.5

- 33 如图是甲、乙两名同学进入高中以来5次体育测试成绩的茎叶图,则甲5次测试成绩的平均数是 \_\_\_\_\_,乙5次测试成绩的平均数与中位数之差是 \_\_\_\_\_.

| 甲 |   |   | 乙 |       |
|---|---|---|---|-------|
| 6 |   | 7 |   | 9     |
| 7 | 4 | 3 | 8 | 0 2 8 |
| 0 |   | 9 |   | 1     |

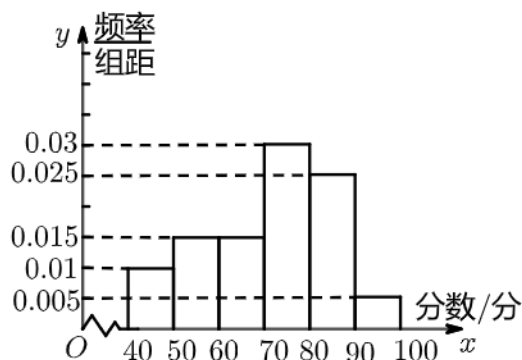
- 34 若数据 $a_1, a_2, \dots, a_{10}$ 的平均数为 $\bar{x}$ ,方差为0.6,则 $a_1, a_2, \dots, a_{10}, \bar{x}$ 这11个数据的方差为 \_\_\_\_\_.

- 35 如图:样本A和B分别取自两个不同的总体,他们的样本平均数分别为 $\bar{x}_A$ 和 $\bar{x}_B$ ,样本标准差分别为 $s_A$ 和 $s_B$ ,则( ).

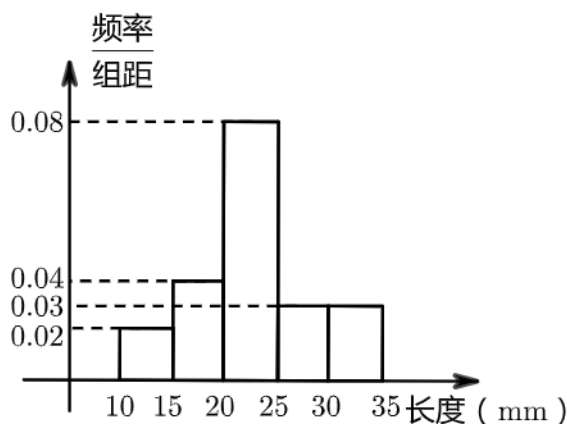


- A.  $\bar{x}_A > \bar{x}_B, s_A > s_B$   
 B.  $\bar{x}_A < \bar{x}_B, s_A > s_B$   
 C.  $\bar{x}_A > \bar{x}_B, s_A < s_B$   
 D.  $\bar{x}_A < \bar{x}_B, s_A < s_B$

- 36 某校从参加高一年级期末考试的学生中抽出60名学生，将其成绩（均为整数）分成六段 $[40, 50)$ ， $[50, 60)$ ， $\dots$ ， $[90, 100]$ 后画出如下频率分布直方图．估计这次考试的平均分为 \_\_\_\_\_ ．

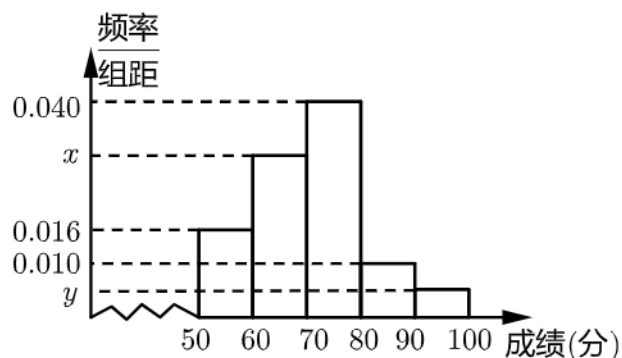


- 37 某工厂对一批新产品的长度（单位： $\text{mm}$ ）进行检测，如下图是检测结果的频率分布直方图，据此估计这批产品的中位数与众数分别为（ ）．



- A. 20 , 22.5                      B. 22.5 , 25  
C. 22.5 , 22.5                      D. 22.75 , 22.75

- 38 为了了解某次考试学生的成绩情况，从中抽取了部分学生的分数（得分取正整数，满分为100）作为样本（样本容量为 $n$ ）进行统计，按照 $[50, 60)$ ， $[60, 70)$ ， $[70, 80)$ ， $[80, 90)$ ， $[90, 100]$ 的分组作出频率分布直方图，已知得分在 $[50, 60)$ ， $[90, 100]$ 的频数分别为8，2．



- (1) 求样本容量 $n$ 和频率分布直方图中的 $x, y$ 的值.
- (2) 估计本次考试学生成绩的众数、中位数、平均数.

## 5. 课堂巩固

- 39 从有300人参加的某项运动的达标测试中, 通过简单随机抽样抽取30人的成绩, 统计成如下表格, 则这300人成绩的平均数的估计值是 \_\_\_\_\_.

|    |   |   |    |   |   |
|----|---|---|----|---|---|
| 分数 | 5 | 4 | 3  | 2 | 1 |
| 人数 | 3 | 9 | 12 | 3 | 3 |

- 40 如图所示的茎叶图记录了一组数据, 关于这组数据, 有下列说法: ①众数是9; ②平均数是10; ③中位数是9或10; ④标准差是3.4. 其中正确的序号是 \_\_\_\_\_.

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 0 | 7 | 8 | 9 | 9 | 9 |
| 1 | 0 | 1 | 2 | 2 | 3 |

- 41 甲、乙两名射击运动员分别进行了5次射击训练, 成绩(单位: 环)如下:

甲: 7, 8, 8, 8, 9

乙: 6, 6, 7, 7, 10;

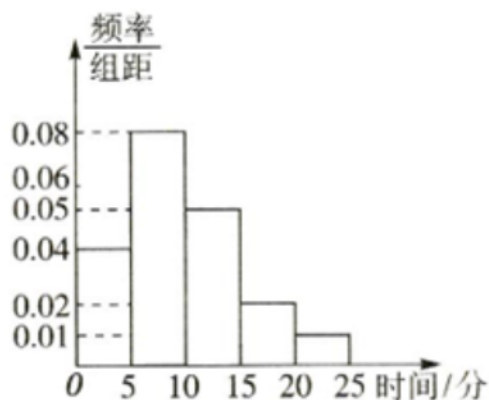
若甲、乙两名运动员的平均成绩分别用 $\bar{x}_1, \bar{x}_2$ 表示, 方差分别为 $s_1^2, s_2^2$ 表示, 则( ).

- A.  $\bar{x}_1 > \bar{x}_2, s_1^2 > s_2^2$
- B.  $\bar{x}_1 > \bar{x}_2, s_1^2 < s_2^2$
- C.  $\bar{x}_1 < \bar{x}_2, s_1^2 < s_2^2$

D.  $\bar{x}_1 < \bar{x}_2$ ,  $s_1^2 > s_2^2$

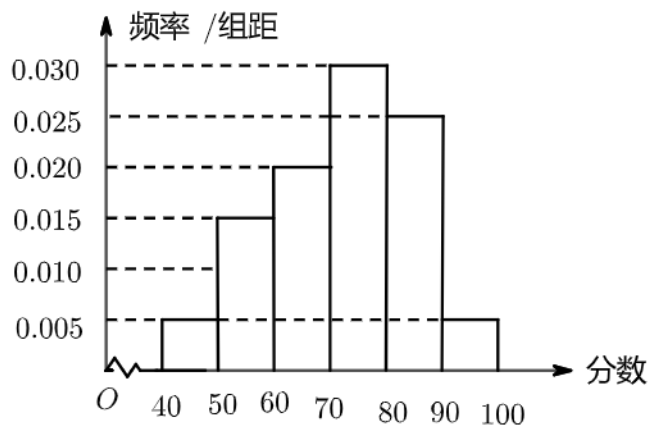
- 42 某医院急诊中心抽查了50名病人，对其等待急诊的时间进行记录，绘制频率分布直方图（每组含

最小值，不含最大值，最后一组含最大值）如图所示：



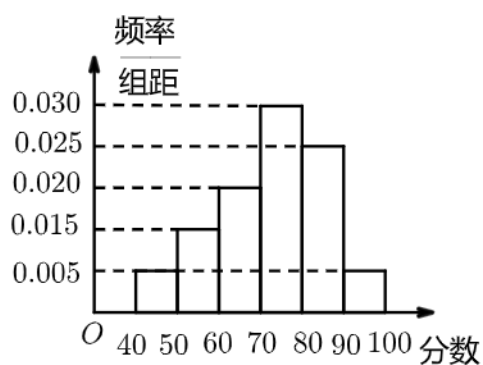
根据频率分布直方图，计算病人平均等待的时间的估计值 $\bar{x}$  = \_\_\_\_\_ 分。

- 43 某校从高一年级期末考试的学生中抽出60名学生，统计其成绩（均为整数）的频率分布直方图如图所示，由此估计此次考试成绩的中位数，众数分别是（ ）。

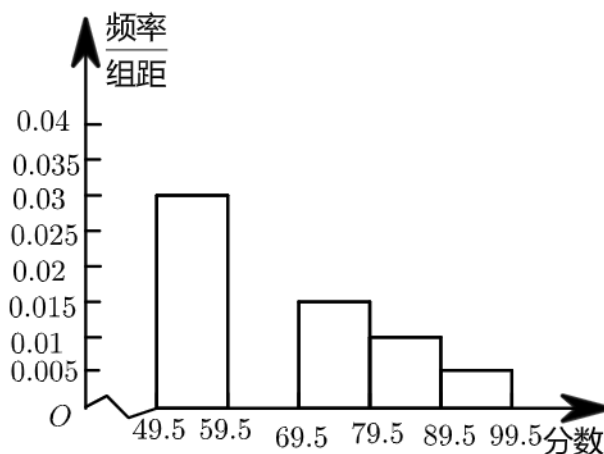


- A. 73.3, 75  
B. 73.3, 80  
C. 70, 70  
D. 70, 75

- 44 某校从参加高二年级学业水平测试的学生中抽出80名学生，其数学成绩（均为整数）的频率分布直方图，如图，估计这次测试中数学成绩的平均分约为 \_\_\_\_\_、众数约为 \_\_\_\_\_、中位数约为 \_\_\_\_\_（结果不能整除的精确到0.1）。



- 45 在某中学举行的电脑知识竞赛中，将高一年级两个班参赛的学生成绩进行整理后分成五组，绘制如图所示的频率分布直方图．已知图中从左到右的第一，第三，第四，第五小组的频率分别是 0.30，0.15，0.10，0.05，第二小组的频数是40．



- (1) 补齐图中频率分布直方图，并求这两个班参赛学生的总人数．
- (2) 利用频率分布直方图，估算本次比赛学生成绩的平均数和中位数．