



第九章 统计

DI JIU ZHANG

9.1 随机抽样



9.1.1 简单随机抽样

新课标新学法 (教师独具内容)

课程标准:1. 通过实例,了解简单随机抽样的含义及其解决问题的过程. 2. 掌握两种简单随机抽样的方法:抽签法和随机数法. 3. 会计算样本均值,了解样本与总体的关系.

教学重点:1. 了解简单随机抽样的概念及其特点. 2. 用样本均值估计总体均值的意义.

教学难点:能合理地从实际问题的总体中抽取样本.

核心概念掌握



知识导学

知识点一 全面调查与抽样调查及相关概念

1. 全面调查

01 对每一个调查对象都进行调查的方法,称为全面调查,又称02 普查.

2. 总体、个体

(1)我们把03 调查对象的全体称为总体.

(2)04 组成总体的每一个调查对象称为个体.

3. 抽样调查

根据一定目的,从总体中05 抽取一部分个体进行调查,并以此为依据06 对总体的情况作出估计和推断的调查方法,称为抽样调查.

4. 样本、样本量、样本数据

(1)我们把07 从总体中抽取的那部分个体称为样本.

(2)08 样本中包含的个体数称为样本量.

(3)09 调查样本获得的变量值称为样本的观测数据,简称10 样本数据.

知识点二 简单随机抽样的定义

1. 放回简单随机抽样和不放回简单随机抽样

一般地, 设一个总体含有 N (N 为正整数) 个个体, 从中逐个抽取 n ($1 \leq n < N$) 个个体作为样本, 如果抽取是 01 放回的, 且每次抽取时总体内的各个个体被抽到的概率 02 都相等, 我们把这样的抽样方法叫做放回简单随机抽样; 如果抽取是 03 不放回的, 且每次抽取时总体内 04 未进入样本的各个个体被抽到的概率 05 都相等, 我们把这样的抽样方法叫做不放回简单随机抽样. 放回简单随机抽样和不放回简单随机抽样统称为 06 简单随机抽样.

2. 简单随机样本

通过 07 简单随机抽样获得的样本称为简单随机样本.

知识点三 常用的简单随机抽样的方法

1. 常用的简单随机抽样的方法

(1) 01 抽签法;

(2) 02 随机数法.

2. 随机数的生成

(1) 用随机试验生成随机数;

(2) 用信息技术生成随机数.

知识点四 总体均值与样本均值

1. 总体均值

一般地, 总体中有 N 个个体, 它们的变量值分别为 $Y_1,$

$Y_2, \dots, Y_N$, 则称 01 $\bar{Y} = \frac{Y_1 + Y_2 + \dots + Y_N}{N} = \frac{1}{N} \sum_{i=1}^N Y_i$ 为总

体均值, 又称总体平均数. 如果总体的 N 个变量值中, 不同的值共有 k ($k \leq N$) 个, 不妨记为 $Y_1, Y_2, \dots, Y_k$, 其中 Y_i 出现的频数 f_i ($i = 1, 2, \dots, k$), 则总体均值还可以写

成加权平均数的形式 02 $\bar{Y} = \frac{1}{N} \sum_{i=1}^k f_i Y_i$.

2. 样本均值

如果从总体中抽取一个容量为 n 的样本, 它们的变量值

分别为 $y_1, y_2, \dots, y_n$, 则称 $\boxed{03} \bar{y} = \frac{y_1 + y_2 + \dots + y_n}{n} =$

$\boxed{04} \frac{1}{n} \sum_{i=1}^n y_i$ 为样本均值, 又称样本平均数.

3. 样本均值与总体均值的关系

我们常用样本均值 $\bar{y}$ $\boxed{05}$ 估计总体均值 $\bar{Y}$.



新知拓展

1. 抽签法的优缺点与操作步骤

(1)优点: 简单易行. 当总体的个数不多时, 使总体处于“搅拌均匀”的状态比较容易, 这时, 每个个体都有均等的机会被抽中, 从而能够保证样本的代表性.

(2)缺点: 仅适用于个体数较少的总体. 当总体容量非常大时, 费时费力又不方便, 况且, 如果号签搅拌的不均匀, 可能导致抽样不公平.

(3)用抽签法从容量为 N 的总体中抽取一个容量为 n 的样本的步骤:

①编号: 给总体中的所有个体编号(号码可以从 1 到 N);

②制作号签: 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);

③均匀搅拌: 将号签放在一个不透明的容器里, 搅拌均匀;

④抽取号码: 每次从容器中不放回地抽取一个号签, 连续抽取 n 次;

⑤构成样本: 从总体中将与抽到的号签上的号码一致的个体抽取, 就构成了一个容量为 n 的样本.

2. 随机数表法的优缺点及操作步骤

(1)优点: 简单易行. 它很好地解决了当总体中的个体数较多时抽签法制签难的问题.

(2)缺点: 当总体中的个体数很多, 需要的样本量也较大时, 用随机数表法抽取样本仍不方便.

(3)随机数表法抽取样本的步骤:

①编号: 对总体的个体进行编号(每个号码位数一致);

②选定初始值: 在随机数表中任选一个数作为开始;

③选号: 从选定的数开始按一定的方向读下去, 得到的号码若不在编号中, 则跳过, 若在编号中, 则取出, 如果得到的号码前面已经取出, 也跳过, 如此继续下去, 直到取满为止;

④确定样本：根据选定的号码抽取样本.

3. 抽签法与随机数法的区别

抽签法适用于总体中个体数较少，样本量也较小的抽样，随机数法适用于总体中个体数较多，但样本量较小的抽样.

4. 用样本估计总体，主要基于以下两点：

一是在很多情况下总体的个数往往很多，甚至无限，不能一一加以考察；二是有些从总体中抽取个体的试验常有破坏性，因而抽取的个体不允许太多。

评价自测

1. 判一判(正确的打“√”，错误的打“×”)

- (1)简单随机抽样就是随便抽取样本. ()
- (2)使用抽签法抽签时，后抽签的人占优势. ()
- (3)利用计算器生成随机数时，按一次“=”键可生成一个随机数. ()

答案 (1)× (2)× (3)√

2. 做一做

(1)下列调查：①每隔5年进行一次人口普查；②报社等进行舆论调查；③灯泡使用寿命的调查；④对入学报名者的学历检查；⑤从20台电视机中抽出3台进行质量检查，其中属于抽样调查的是()

- A. ①②③
- B. ②③⑤
- C. ②③④
- D. ①③⑤

(2)下列抽样试验中，适合用抽签法的有()

- A. 从某厂生产的3000件产品中抽取600件进行质量检验
- B. 从某厂生产的两箱(每箱15件)产品中抽取6件进行质量检验
- C. 从甲、乙两厂生产的两箱(每箱15件)产品中抽取6件进行质量检验
- D. 从某厂生产的3000件产品中抽取10件进行质量检验

答案 (1)B (2)B

核心素养形成

题型一 简单随机抽样的判断

例1 下列5个抽样中，简单随机抽样的个数是()

- ①从无数个个体中抽取50个个体作为样本；
- ②仓库中有1万支火炬，从中一次性抽取100支火炬进行质量检查；
- ③某连队从200名党员官兵中，挑选出50名最优秀的官兵赶赴青海参加抗震救灾工作；

④一彩民选号，从装有 36 个大小、形状都相同的号签的盒子中无放回地抽出 6 个号签.

A. 0

B. 1

C. 2

D. 3

【解析】 根据简单随机抽样的特点逐个判断. ①不是简单随机抽样，因为简单随机抽样要求被抽取样本的总体中的个体数是有限的；②不是简单随机抽样，虽然“一次性抽取”和“逐个抽取”不影响个体被抽到的可能性，但简单随机抽样要求的是“逐个抽取”；③不是简单随机抽样，因为 50 名官兵是从中挑出来的，是最优秀的，每个个体被抽到的可能性不同，不符合简单随机抽样中“等可能抽样”的要求；④是简单随机抽样，因为总体中的个体数是有限的，并且是从总体中逐个进行抽取的，是不放回、等可能的抽样. 综上，只有④是简单随机抽样.

【答案】 B

金版点睛 简单随机抽样必须具备的特点

(1)被抽取样本的总体中的个体数是有限的.

(2)抽取的样本是从总体中逐个抽取的.

(3)简单随机抽样是一种等可能的抽样.

如果三个特征有一个不满足，就不是简单随机抽样.

跟踪训练1

判断下面的抽样方法是否为简单随机抽样，并说明理由.

(1)某班 45 名同学，指定个子最矮的 5 名同学参加学校组织的某项活动；

(2)从 20 个零件中一次性抽出 3 个进行质量检查.

解 (1)不是简单随机抽样.

因为指定个子最矮的 5 名同学，是在 45 名同学中特指的，不存在随机性，不是等可能抽样.

(2)不是简单随机抽样.

因为一次性抽取 3 个不是逐个抽取，不符合简单随机抽样的特征.

题型二 用抽签法抽取样本

例 2 (1)上海某中学从 40 名学生中选 1 人作为上海男篮啦啦队的成员，采用下面两种选法，则抽签法的序号是_____.

①将这 40 名学生从 1~40 进行编号,相应地制作 1~40 的 40 个号签,把这 40 个号签放在一个暗箱中搅匀,然后随机地从中抽取 1 个号签,与这个号签编号一致的学生幸运入选;

②将 39 个白球与 1 个红球(球除颜色外,其他完全相同)混合放在一个暗箱中搅匀,让 40 名学生逐一从中摸取一球,摸到红球的学生成为啦啦队成员.

(2)在社区公益活动中,某单位共有 50 名志愿者参与了报名,现要从中随机抽出 6 人参加一项活动,请用抽签法进行抽样,并写出过程.

[解析] (1)①满足抽签法的特征,是抽签法;②不是抽签法,因为抽签法要求所有的号签编号互不相同,而②中 39 个白球无法相互区分.

(2)第一步,将 50 名志愿者编号,号码依次为 1,2,3, ..., 50;

第二步,将号码分别写在大小、形状、质地都相同的纸条上,揉成团,制成号签;

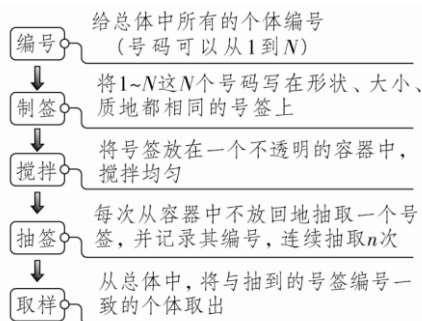
第三步,将所有号签放入一个不透明的箱子中,搅拌均匀;

第四步,一次取出 1 个号签,连取 6 次(不放回抽取),并记录其编号;

第五步,将对应编号的志愿者选出即可.

[答案] (1)① (2)见解析

金版点睛 抽签法的五个步骤



跟踪训练2

从 20 架钢琴中抽取 5 架进行质量检查,请用抽签法确定这 5 架钢琴.

解 第一步,将 20 架钢琴编号,号码是 01,02, ..., 20.

第二步,将号码分别写在相同的纸条上,揉成团,制成号签.

第三步,将得到的号签放入一个不透明的袋子中,并充分搅匀.

第四步,从袋子中逐个不放回地抽取 5 个号签,并记录上面的编号.

第五步,所得号码对应的 5 架钢琴就是要进行质量检查的对象.

题型三 用随机数法抽取样本

例 3 (1)要考察某种品牌的 850 颗种子的发芽率,从中抽取 50 颗种子进行实验,利用随机数表法抽取种子,先将 850 颗种子按 001,002, ..., 850 进行编号,如果从随机数表第 3 行第 6 列的数开始向右读,请依次写出最先检验的 4 颗种子的编号: _____.(下面抽取了随机数表第 1 行至第 8 行)

03 47 43 73 86	36 96 47 36 61	46 98 63 71 62	33 26 16 80 45	60 11 14 10 95
97 74 24 67 62	42 81 14 57 20	42 53 32 37 32	27 07 36 07 51	24 51 79 89 73
16 76 62 27 66	56 50 26 71 07	32 90 79 78 53	13 55 38 58 59	88 97 54 14 10
12 56 85 99 26	96 96 68 27 31	05 03 72 93 15	57 12 10 14 21	88 26 49 81 76
55 59 56 35 64	38 54 82 46 22	31 62 43 09 90	06 18 44 32 53	23 83 01 30 30
16 22 77 94 39	49 54 43 54 82	17 37 93 23 78	87 35 20 96 43	84 26 34 91 64
84 42 17 53 31	57 24 55 06 88	77 04 74 47 67	21 76 33 50 25	83 92 12 06 76
63 01 63 78 59	16 95 55 67 19	98 10 50 71 75	12 86 73 58 07	44 39 52 38 79

(2)现有一批零件,其编号为 600,601,602, ..., 999.利用原有的编号从中抽取一个容量为 10 的样本进行质量检查,若用信息技术生成随机数法,怎样设计方案?

[解析] (1)从随机数表第 3 行第 6 列的数 2 开始向右读,第一个小于 850 的数字是 227,第二个数字是 665,第三个数字是 650,第四个数字是 267,符合题意.

(2)用计算器生成随机数,第一步,进入计算器的计算模式,调出生成随机数的函数并设置参数;

第二步,按“=”键生成一个符合条件的随机数,继续重复按“=”键,生成多个随机数,如果生成的随机数重复,则跳过去不读,直到产生 10 个没有重复的随机数为止;

第三步,以上 10 个号码对应的 10 个零件就是要抽取的对象.(答案不唯一)

[答案] (1)227,665,650,267 (2)见解析

金版点睛 利用随机数表法抽样时应注意的问题

(1)编号要求位数相同,若不相同需先调整到一致后再进行抽样,如当总体中有 100 个个体时,为了操作简便可以选择从 00 开始编号,那么所有个体的号码都用两位数字表示即可,从 00~99 号.如果选择从 001 开始编号那么所有个体的号

码都必须用三位数字表示,即从 001 ~ 100.很明显每次读两个数字要比读三个数字节省时间.

(2)第一个数字的抽取是随机的.

(3)当随机数选定,开始读数时,读数的方向可左、可右、可上、可下,但应是事先定好的.

(4)读数不在总体编号内的和已取出的不算,依次下去,直至得到容量为 n 的样本.

跟踪训练3

(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: 从随机数表第 1 行的第 5 列和第 6 列数字开始由左到右一次选取两个数字, 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)某合资企业有 3000 名职工,要从中随机抽出 200 人去参观学习. 请用信息技术生成随机数法进行抽取, 并写出过程.

答案 (1)D (2)见解析

解析 (1)从随机数表第 1 行的第 5 列和第 6 列数字开始由左到右一次选取两个数字开始向右读, 第一个数为 65, 不符合条件, 第二个数为 72, 不符合条件, 第三个数为 08, 符合条件, 以下符合条件的数依次为 02,14,07,01, 故第 5 个数为 01.故选 D.

(2)第一步, 将 3000 名职工依次编号为 1,2,3, ..., 3000;

第二步, 用电子表格软件生成随机数, 在电子表格软件的任一单元格中, 输入 “=RANDBETWEEN(1,3000)”, 则生成一个 1 ~ 3000 范围内的整数随机数;

第三步, 利用电子表格软件的自动填充功能得到 200 个没有重复的随机数;

第四步, 这 200 个号码对应的 200 名职工就是要抽取的职工.

题型四 用样本均值估计总体均值

例 4 某灯泡厂为测量一批灯泡的使用寿命, 从中随机抽查了 20 只灯泡, 它

们的使用寿命变量值(单位: h)如下所示:

624 847 1205 698 1845 2457 618 1325 1908 2426
2018 2248 2465 2576 987 737 1628 1998 2543 2007

则由这些样本观测数据，估计这批灯泡的平均使用寿命是多少？

[解] 抽出的 20 只灯泡的使用寿命组成一个样本, 可以用样本的平均使用寿命来估计这批灯泡的平均使用寿命.

根据题中数据, 可得样本的均值为 1658 h.

因此, 可以估计这批灯泡的平均使用寿命大约是 1658 h.

金版点睛

- (1) 计算数据的加权平均数, 需理解组中值的意义和数据“权数”的意义.
- (2) 用样本的平均数估计总体的平均数, 体现了重要的统计思想.

跟踪训练4

为了解一批轮胎的性能,汽车制造厂从这批轮胎中随机抽取了 8 个进行测试,每个轮胎行驶的最远里程数(单位: 1000 km)为: 96,112,97,108,100,103,86,98.则估计这批轮胎行驶的最远里程数的平均数为()

- A. 100
B. 99
C. 98
D. 97

答案 A

解析 用样本平均数估计总体平均数，得这批轮胎行驶的最远里程数的平均数约为

$$\frac{96 + 112 + 97 + 108 + 100 + 103 + 86 + 98}{8} = 100.$$

随堂水平达标

1. 为了检查一批光盘的质量，从中抽取了 500 张进行检测，则这个问题中样本量是()

- A. 500 张光盘
B. 500
C. 500 张光盘的质量
D. 这批光盘

答案 B

解析 样本中包含的个体数称为样本量, 故这个问题中样本量是 500. 故选 B.

2. 下列抽样方法是简单随机抽样的是()

- A. 从 100 个学生家长中一次性随机抽取 10 人做家访
- B. 从 38 本教辅参考资料中选取内容讲解较好的 3 本作为教学参考
- C. 从自然数集中一次性抽取 20 个进行奇偶性分析
- D. 某参会人员从最后一排 20 个座位中随机选择一个坐下

答案 D

解析 A 不是简单随机抽样, 因为是“一次性”抽取; B 不是简单随机抽样, 因为每个个体被抽到的概率不相等; C 不是简单随机抽样, 因为是“一次性”抽取, 且“总体容量无限”. D 是简单随机抽样.

3. 从 52 名学生中选取 5 名学生参加“希望杯”全国数学邀请赛, 若采用简单随机抽样抽取, 则每人入选的可能性()

- A. 都相等, 且为 $\frac{1}{52}$
- B. 都相等, 且为 $\frac{1}{10}$
- C. 都相等, 且为 $\frac{5}{52}$
- D. 都不相等

答案 C

解析 对于简单随机抽样, 在抽样过程中每一个个体被抽取的机会都相等(随机抽样的等可能性). 若样本容量为 n , 总体的个体数为 N , 则用简单随机抽样时, 每一个个体被抽到的可能性都是 $\frac{n}{N}$, 体现了这种抽样方法的客观性和公平性. 因此每人入选的可能性都相等, 且为 $\frac{5}{52}$.

4. 从某批零件中抽取 50 个, 然后再从 50 个中抽出 40 个进行合格检查, 发现合格品有 36 个, 则该产品的合格率约为()

- A. 36%
- B. 72%
- C. 90%
- D. 25%

答案 C

解析 $\frac{36}{40} \times 100\% = 90\%$.

5. 为了调查某班学生每天完成家庭作业所需的时间, 在该班随机抽查了 8 名学生, 他们每天完成家庭作业所需时间(单位: 分钟)分别为 60, 55, 75, 55, 55, 43, 65, 40.

(1) 求这组样本观测数据的平均数;

(2) 估计该班学生每天完成家庭作业的平均时间, 按照学校要求, 学生每天完

成家庭作业所需的平均时间不能超过 60 分钟,该班学生每天完成家庭作业的平均时间是否符合学校的要求?

解 (1)这组样本观测数据的平均数为 $\frac{1}{8} \times (60 + 55 + 75 + 55 + 55 + 43 + 65 + 40)$
 $= 56.$

(2)由样本平均数,估计该班学生每天完成家庭作业的平均时间为 56 分钟.

$\because 56 < 60, \therefore$ 该班学生每天完成家庭作业的平均时间符合学校的要求.



9.1.2 分层随机抽样



9.1.3 获取数据的途径

核心概念掌握



知识导学

知识点一 分层随机抽样的定义

一般地,按一个或多个变量把总体划分成01若干个子总体,每个个体属于且仅属于一个子总体,在每个子总体中独立地进行02简单随机抽样,再把所有子总体中抽取的样本合在一起作为03总样本,这样的抽样方法称为04分层随机抽样.

(1)层: 每一个05子总体称为层.

(2)比例分配: 在分层随机抽样中,如果每层样本量都与层的大小成06比例,那么称这种样本量的分配方式为07比例分配.

知识点二 分层随机抽样的步骤

(1)分层: 按某种特征将总体分成若干部分,分层需遵循01不重复、不遗漏的原则.

(2)按比例确定每层抽取个体的个数,抽取比例由每层个体占02总体的比例确定.

(3)各层分别按03简单随机抽样的方法抽取.

(4)综合每层抽样,组成样本.

知识点三 样本平均数估计总体平均数

1. 在分层随机抽样中,如果层数分为2层,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 .我们用 $X_1, X_2, \dots, X_M$ 表示第1层各个个体的变量值,用 $x_1, x_2, \dots, x_m$ 表示第1层样本的各个个体的变量值;用 $Y_1, Y_2, \dots, Y_N$ 表示第2层各个个体的变量值,用 $y_1, y_2, \dots, y_n$ 表示第2层样本的各个个体的变量值,则各层个体及样本特征可由下表表示出来:

项目 层数	个 体 数	样 本 量	总体平均数	样本平均数
第 1 层	M	m	$\bar{X} =$ $\text{[01]} \frac{X_1 + X_2 + \cdots + X_M}{M}$ $= \frac{1}{M} \sum_{i=1}^M x_i$	$\bar{x} = \frac{x_1 + x_2 + \cdots + x_m}{m}$ $= \frac{1}{m} \sum_{i=1}^m x_i$
第 2 层	N	n	$\bar{Y} =$ $\text{[02]} \frac{Y_1 + Y_2 + \cdots + Y_N}{N}$ $= \frac{1}{N} \sum_{i=1}^N Y_i$	$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}$ $= \frac{1}{n} \sum_{i=1}^n y_i$

2. 总体平均数 $\bar{W} = \text{[03]} \frac{\sum_{i=1}^M X_i + \sum_{i=1}^N Y_i}{M + N}$, 样本平均数 $\bar{w} =$

$$\text{[04]} \frac{\sum_{i=1}^m x_i + \sum_{i=1}^n y_i}{m + n}.$$

3. 总体平均数 $\bar{W}$ 可由样本平均数 $\bar{w}$ 来估计.

$$\bar{W} = \text{[05]} \frac{M}{M+N} \bar{x} + \frac{N}{M+N} \bar{y} = \text{[06]} \frac{m}{m+n} \bar{x} + \frac{n}{m+n} \bar{y} = \bar{w}.$$

知识点四 获取数据的途径

统计学是通过 [01] 收集数据 和 [02] 分析数据 来认识未知现象的. 获取数据的一些基本途径有: [03] 通过调查获取数据、[04] 通过试验获取数据、[05] 通过观察获取数据、[06] 通过查询获取数据.



新知拓展

1. 分层随机抽样的特点

由于层与层之间有明显的区别, 而层内个体间的差异不明显, 在抽取样本时, 一般在每层抽样时随机抽样, 以保证公平、公正, 在实际操作时, 应先计算抽样

比 $k = \frac{\text{样本量}}{\text{总体个数}}$ ，然后再计算每层需要抽取的个数：抽样比 $\times$ 该层个体数目 = $\frac{\text{样本量}}{\text{总体个数}} \times \text{该层个体数目}$ 。

2. 分层随机抽样的适用条件

分层随机抽样尽量利用事先所掌握的各种信息，并充分考虑保持样本结构与总体结构的一致性，这对提高样本的代表性非常重要。当总体是由差异明显的几个部分组成时，往往选用分层随机抽样的方法。

☒ 评价自测

1. 判一判(正确的打“√”，错误的打“×”)

- (1)在分层随机抽样中，每层被抽到的个体数是一样的。()
- (2)分层随机抽样中每个个体被抽到的可能性不一样。()
- (3)样本平均数一定大于总体平均数。()

答案 (1)× (2)× (3)×

2. 做一做

(1)为调查某班学生的平均身高，从 50 名学生中抽取 5 名，因为男生的身高和女生的身高有显著不同，所以获取样本时宜采用_____抽样。

(2)一个班共有 54 人，其中男女人数比为 5:4，若抽取 9 人参加教改调查会，则应抽取男同学_____人。

(3)已知数据 x_1, x_2, x_3, x_4 的平均数为 2，则数据 $3x_1, 3x_2, 3x_3, 3x_4$ 的平均数是_____。

答案 (1)分层随机 (2)5 (3)6

核心素养形成

题型一 分层随机抽样的概念

例 1 (1)某政府机关在编人员共 100 人，其中副处级以上干部 10 人，一般干部 70 人，工人 20 人，上级部门为了了解该机关对政府机构改革的意见，要从中抽取 20 人，用下列哪种方法最合适()

- A. 抽签法
- B. 简单随机抽样法
- C. 分层随机抽样法
- D. 随机数法

(2)分层随机抽样又称类型随机抽样，即将相似的个体归入一类(层)，然后每类抽取若干个个体构成样本，若分层随机抽样要保证每个个体等可能抽样，必须进行()

- A. 每层等可能抽样

- B. 每层可以不等可能抽样
- C. 所有层按同一抽样比等可能抽样
- D. 所有层抽取个体数量相同

[解析] (1)总体由差异明显的三部分构成, 应选用分层随机抽样法.

(2)保证每个个体等可能的被抽取是简单随机抽样和比例分配的分层随机抽样的共同特征, 为了保证这一点, 分层随机抽样时必须在所有层都按同一抽样比等可能抽取.

[答案] (1)C (2)C

金版点睛 使用分层随机抽样的原则

(1)将相似的个体归入一类, 即为一层, 分层要求每层的各个个体互不交叉, 即遵循不重复、不遗漏的原则.

(2)比例分配的分层随机抽样为保证每个个体等可能入样, 需遵循在各层中进行简单随机抽样, 使每层样本数量与每层个体数量的比等于抽样比.

跟踪训练1

下列问题中, 最适合用分层随机抽样方法抽样的是()

- A. 某电影院有 32 排座位, 每排有 40 个座位, 座位号是 1~40.有一次报告会坐满了听众, 报告会结束以后为听取意见, 要留下 32 名听众进行座谈
- B. 从 10 台冰箱中抽出 3 台进行质量检查
- C. 某乡农田有山地 8000 亩, 丘陵 12000 亩, 平地 24000 亩, 洼地 4000 亩, 现抽取农田 480 亩估计全乡农田平均产量
- D. 从 50 个零件中抽取 5 个做质量检验

答案 C

解析 A 的个体没有呈现出较大差异, 不适合用分层随机抽样法; B 的总体容量较小, 用简单随机抽样法比较方便; C 总体容量较大, 且各类田地的产量差别很大, 宜采用分层随机抽样方法; D 与 B 类似.

题型二 分层随机抽样的应用

例 2 一个单位有职工 500 人, 其中不到 35 岁的有 125 人, 35 岁至 49 岁的有 280 人, 50 岁及 50 岁以上的有 95 人, 为了了解这个单位职工与身体状态有关的某项指标, 要从中抽取 100 名职工作为样本, 若职工年龄与这项指标有关, 应该怎样抽取?

[解] 用比例分配的分层随机抽样来抽取样本, 步骤如下:

(1)分层. 按年龄将 500 名职工分成三层: 不到 35 岁的职工; 35 岁至 49 岁的职工; 50 岁及 50 岁以上的职工.

(2)确定每层抽取个体的个数. 抽样比为 $\frac{100}{500} = \frac{1}{5}$, 则在不到 35 岁的职工中抽取 $125 \times \frac{1}{5} = 25$ (人);

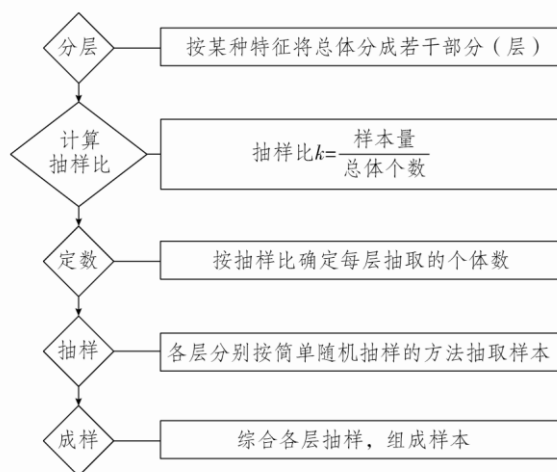
在 35 岁至 49 岁的职工中抽取 $280 \times \frac{1}{5} = 56$ (人);

在 50 岁及 50 岁以上的职工中抽取 $95 \times \frac{1}{5} = 19$ (人).

(3)在各层分别按随机数法抽取样本.

(4)汇总每层抽样, 组成样本.

金版点睛 利用比例分配的分层随机抽样抽取样本的操作步骤



跟踪训练2

某电视台在因特网上就观众对某一节目的喜爱程度进行调查, 参加调查的总人数为 12000 人, 其中持各种态度的人数如下表所示:

很喜爱	喜爱	一般	不喜爱
2435	4567	3926	1072

电视台为了进一步了解观众的具体想法和意见, 打算从中再抽取 60 人进行更为详细的调查, 应怎样进行抽样?

解 采用比例分配的分层随机抽样的方法, 抽样比为 $\frac{60}{12000} = \frac{1}{200}$.

“很喜爱”的有 2435 人, 应抽取 $2435 \times \frac{1}{200} \approx 12$ (人);

“喜爱”的有 4567 人, 应抽取 $4567 \times \frac{1}{200} \approx 23$ (人);

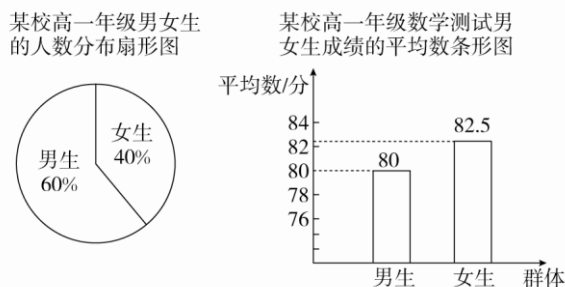
“一般”的有 3926 人, 应抽取 $3926 \times \frac{1}{200} \approx 20(\text{人})$;

“不喜爱”的有 1072 人, 应抽取 $1072 \times \frac{1}{200} \approx 5(\text{人})$.

因此, 采用分层随机抽样的方法在“很喜爱”“喜爱”“一般”“不喜爱”的人中应分别抽取 12 人、23 人、20 人、5 人.

题型三 分层随机抽样的综合应用

例 3 某校高一年级有 24 个班, 共 1000 名学生, 他们参加了一次数学测试. 学校统计了所有学生的成绩, 得到下列统计图.



(1) 求该校高一年级学生本次测试成绩的平均数.

(2) 假设要抽取 300 名学生, 按照比例分配的分层随机抽样的方法, 应抽取男生多少名? 女生多少名?

[解] (1) 由题意并结合扇形统计图, 可知男生共有 $1000 \times 60\% = 600(\text{名})$, 女生共有 $1000 \times 40\% = 400(\text{名})$. 由成绩的平均数条形图可得, 该校高一年级学生本次测试成绩的平均数 $\bar{x} = (80 \times 600 + 82.5 \times 400) \div 1000 = 81(\text{分})$.

(2) 抽取 300 名学生, 采用比例分配的分层随机抽样的方法, 则男生应抽取 $600 \times \frac{300}{1000} = 180(\text{名})$, 女生应抽取 $400 \times \frac{300}{1000} = 120(\text{名})$.

金版点睛

(1) 解决统计问题要做好数据的处理和分析.

(2) 条形统计图能清楚地表示出每个项目的数据, 扇形统计图直接反映部分在总体的百分比大小.

跟踪训练3

有 4 万个不小于 70 的两位数, 从中随机抽取了 3000 个数, 统计如下表:

数据 x	$70 < x < 79$	$80 < x < 89$	$90 < x < 99$
个数	800	1300	900
平均数	78.1	85	91.9

请根据表格中的信息，估计这 4 万个数的平均数约为()

- A. 92.16
- B. 85.23
- C. 84.73
- D. 77.97

答案 B

解析 这 3000 个数的平均数为：

$$\frac{78.1 \times 800 + 85 \times 1300 + 91.9 \times 900}{3000} = 85.23,$$

于是用样本的平均数去估计总体的平均数,则这 4 万个数的平均数约为 85.23.

题型四 获取数据的途径

例 4 简单设计一份问卷，调查学生对高一各学科的态度.

[解] 请按自己的感受把下面这些学科的序号填在空格里.

- ①语文 ②数学 ③外语 ④物理 ⑤化学 ⑥生物 ⑦历史 ⑧地理 ⑨
- 政治 ⑩体育 ⑪艺术(音乐、美术) ⑫技术

我喜欢的学科	
我感觉压力最大的学科	
我不喜欢的学科	
我觉得有用的学科	
我觉得内容多的学科	
我觉得内容少的学科	

金版点睛 调查问卷中问题设计的要求

- (1)问卷中的问题必须设计详细，以便被调查者顺利回答.
- (2)把比较容易的，不涉及个人的问题排在比较靠前的位置，较难的、涉及个人的问题放在后面.

跟踪训练4

为调查小区平均每户居民的月用水量，下面是三名同学设计的方案：

学生甲：我把这个用水量调查表放在互联网上，只要登陆网站的人就可以看到这张表，他们填的表可以很快地反馈到我的电脑中，这样就可以很快估算出小区平均每户居民的月用水量；

学生乙：我给我们居民小区的每一个住户发一张用水调查表，只要一两天就可以统计出小区平均每户居民的月用水量；

学生丙：我在小区的电话号码本上随机地选出一定数量的电话号码，然后逐

个给这些住户打电话，问一下他们的月用水量，然后就可以估算出小区平均每户居民的月用水量。

请问：这三位同学设计的方案中哪一个较合理？你有何建议？

解 学生甲的方法得到的样本只能够反映上网居民的用水情况，它是一种方便样本，所得到的样本代表性差，不能很准确地获得平均每户居民的月用水量。

学生乙的方法实际上是普查，花费的人力、物力更多一些，但是如果统计过程不出错，就可以准确地得到平均每户居民的月用水量。

学生丙的方法是一种随机抽样的方法，所在小区的每户居民都装有电话的情况下，建议用随机抽样方法获得数据，用学生丙的方法，既节省人力、物力，又可以得到比较精确的结果。

随堂水平达标

1. 某校高三年级有男生 500 人，女生 400 人，为了解该年级学生的健康状况，从男生中任意抽取 25 人，从女生中任意抽取 20 人进行调查。这种抽样方法是 ()

A. 简单随机抽样

B. 抽签法

C. 随机数法

D. 分层随机抽样

答案 D

解析 从男生 500 人中抽取 25 人，从女生 400 人中抽取 20 人，抽取的比例相同，因此用的是分层随机抽样，且样本量的分配方式为比例分配。

2. 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和比例分配的分层随机抽样两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 ，则 ()

A. $P_1 > P_2$

B. $P_1 < P_2$

C. $P_1 = P_2$

D. 无法确定

答案 C

解析 简单随机抽样和比例分配的分层随机抽样都是等可能抽样，每个个体被抽中的概率均为 $\frac{n}{N}$ 。

3. 甲校有 3600 名学生，乙校有 5400 名学生，丙校有 1800 名学生，为统计三校学生某方面的情况，计划采用比例分配的分层随机抽样法抽取一个容量为 90

的样本，应在这三校分别抽取学生()

- A. 30 人, 30 人, 30 人 B. 30 人, 45 人, 15 人
C. 20 人, 30 人, 10 人 D. 30 人, 50 人, 10 人

答案 B

解析 先求抽样比 $\frac{n}{N} = \frac{90}{3600 + 5400 + 1800} = \frac{1}{120}$, 再各层按抽样比分别抽取,

则甲校抽取 $3600 \times \frac{1}{120} = 30(\text{人})$, 乙校抽取 $5400 \times \frac{1}{120} = 45(\text{人})$, 丙校抽取

$1800 \times \frac{1}{120} = 15(\text{人})$, 故选 B.

4. 某大学为了解在校本科生对参加某项社会实践活动的意向, 拟采用比例分配的分层随机抽样的方法, 从该校四个年级的本科生中抽取一个容量为 300 的样本进行调查, 已知该校一年级、二年级、三年级、四年级的本科生人数之比为 4 : 5 : 5 : 6, 则应从一年级本科生中抽取_____名学生.

答案 60

解析 根据题意, 应从一年级本科生中抽取的人数为 $\frac{4}{4+5+5+6} \times 300 = 60$.

5. 某校 500 名学生中, O 型血有 200 人, A 型血有 125 人, B 型血有 125 人, AB 型血有 50 人, 为了研究血型与色弱的关系, 需从中抽取一个容量为 20 的样本. 按照比例分配的分层随机抽样方法抽取样本, 各种血型的人分别抽多少?

解 用比例分配的分层随机抽样方法抽样.

$$\therefore \frac{20}{500} = \frac{1}{25}, \therefore 200 \times \frac{1}{25} = 8, 125 \times \frac{1}{25} = 5, 50 \times \frac{1}{25} = 2.$$

故 O 型血抽 8 人, A 型血抽 5 人, B 型血抽 5 人, AB 型血抽 2 人.



第九章 统计

DI JIU ZHANG

9.2 用样本估计总体



9.2.1 总体取值规律的估计

新课标新学法 (教师独具内容)

课程标准: 1. 结合实例,能用样本估计总体的取值规律. 2. 能根据实际问题的特点,选择恰当的统计图表对数据进行可视化描述,体会合理使用统计图表的重要性.

教学重点: 掌握频率分布表、频率分布直方图的制作步骤.

教学难点: 能通过样本的频率分布估计总体的频率分布.

核心概念掌握



知识导学

知识点一 频率分布表与频率分布直方图

1. 一般地,当总体很大或不便获取时,用 01 样本的频率分布估计 02 总体的频率分布. 把反应总体频率分布的表格称为 03 频率分布表.

2. 把表示 04 样本数据分布规律的图形称为 05 频率分布直方图.

3. 制作频率分布表与频率分布直方图的一般步骤:

(1)求极差,即计算 06 最大值与最小值的差.

(2)决定 07 组距与组数.

(3)将数据 08 分组.

(4)列 09 频率分布表.

(5)画频率分布直方图.

知识点二 常用的统计图

1. 条形图

条形图能清楚地表示出每个项目的 01 具体数量.

2. 折线图

折线图能够清晰地反映数据的 02 变化趋势或情况.

3. 扇形图

扇形图能直观、生动地反映各部分在总体中所占的 03 比例情况.



新知拓展

1. 频率分布直方图能够直观、形象地反映样本的分布规律,可以大致估计出总体的分布,但是从频率分布直方图中得不出原始的数据内容,把数据绘制成频率分布直方图后,原有的具体数据信息就被抹掉了. 在频率分布直方图中,由于长方形的面积 $S = \text{组距} \times \frac{\text{频率}}{\text{组距}} = \text{频率}$, 所以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在各个小组的频率的大小,各个小长方形的面积总和等于 1.

2. 一般地,样本量越大,这种估计就越精确. 总体估计要掌握: (1) “表”(频率分布表); (2) “图”(频率分布直方图).

提醒: 直方图的纵轴(小长方形的高)一般是频率除以组距的商,横轴一般是数据的大小,小长方形的面积表示频率.

☒ 评价自测

1. 判一判(正确的打“√”,错误的打“×”)

(1) 频率分布直方图的纵轴表示频率. ()

(2) 数据落在各小组内的频率用 $\frac{\text{各小组频数}}{\text{样本量}}$ 来表示. ()

(3) 频率分布直方图把样本数据落在各小组的比例大小直观化,更有利于我们从整体上把握数据分布的特点. ()

答案 (1)× (2)√ (3)√

2. 做一做

(1) 将容量为 100 的样本数据分为 8 个组,如下表:

组号	1	2	3	4	5	6	7	8
频数	10	13	x	14	15	13	12	9

则第 3 组的频率为()

A. 0.03

B. 0.07

C. 0.14

D. 0.21

(2) 一个容量为 n 的样本,分成若干组,已知某组的频率和频数分别为 0.125 和 40,则 n 的值为()

A. 640

B. 320

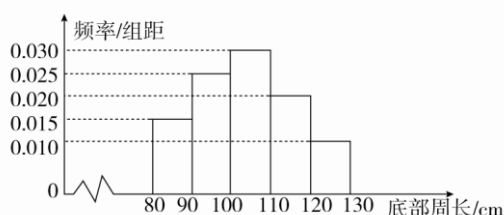
C. 240

D. 160

(3)一个容量为 20 的样本，分组后，组距与频数如下：[10,20)，2；[20,30)，3；[30,40)，4；[40,50)，5；[50,60)，4；[60,70)，2，则样本在[10,50)上的频率为()

A. $\frac{1}{20}$ B. $\frac{1}{4}$ C. $\frac{1}{2}$ D. $\frac{7}{10}$

(4)为了了解一片经济林的生长情况，随机抽测了其中 60 株树木的底部周长(单位：cm)，所得数据均在区间[80,130]上，其频率分布直方图如图所示，则在抽测的 60 株树木中，有_____株树木的底部周长小于 100 cm.



答案 (1)C (2)B (3)D (4)24

核心素养形成

题型一 频率分布直方图的画法

例 1 从某校高一年级 1002 名新生中随机抽取一个容量为 100 的身高样本，数据如下(单位：cm)，试作出该样本的频率分布表和频率分布直方图.

168	165	171	167	170	165	170	152	175	174
165	170	168	169	171	166	164	155	164	158
170	155	166	158	155	160	160	164	156	162
160	170	168	164	174	171	165	179	163	172
180	174	168	164	174	171	165	179	163	169
151	168	158	168	176	155	165	165	169	162
177	158	175	165	169	151	163	166	163	167
178	165	158	170	169	159	155	163	153	155
167	163	164	158	168	167	161	162	167	168
161	165	174	156	167	166	162	161	164	166

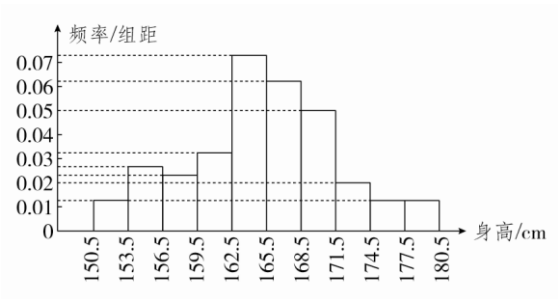
[解] (1)在全部数据中找出最大值 180，最小值 151，计算极差 = 180 - 151 = 29.

(2)取组距为 3，则组数为 $\frac{29}{3}\approx 10$.

(3)从第一组[150.5,153.5)开始，统计出各组的频数，再计算各组的频率，并将结果填入下表：

分 组	频 数 统 计	频 数	频 率
[150.5,153.5)	正	4	0.04
[153.5,156.5)	正 下	8	0.08
[156.5,159.5)	正 丁	7	0.07
[159.5,162.5)	正 正	10	0.10
[162.5,165.5)	正 正 正 正 下	23	0.23
[165.5,168.5)	正 正 正 正	19	0.19
[168.5,171.5)	正 正 正	15	0.15
[171.5,174.5)	正 一	6	0.06
[174.5,177.5)	正	4	0.04
[177.5,180.5]	正	4	0.04
合 计		100	1

频率分布直方图如图.



金版点睛 绘制频率分布直方图的注意事项

- (1)计算极差，需要找出这组数的最大值和最小值，当数据很多时，可选一个数当参照.
- (2)将一批数据分组，目的是要描述数据分布规律，要根据数据多少来确定分组数目，一般来说，数据越多，分组越多.
- (3)将数据分组，决定分点时，一般使分点比数据多一位小数，并且把第一组

的起点稍微减小一点.

(4)列频率分布表时,可通过逐一判断各个数据落在哪个小组内,以“正”字确定各个小组内数据的个数.

(5)画频率分布直方图时,纵坐标表示频率与组距的比值,一定不能标成频率.

 跟踪训练1

美国历届总统中,就任时年龄最小的是罗斯福,他于 1901 年就任,当时年仅 42 岁;就任时年龄最大的是特朗普,他于 2016 年就任,当时 70 岁.下面按时间顺序(从 1789 年的华盛顿到 2016 年的特朗普,共 45 任)给出了历届美国总统就任时的年龄:

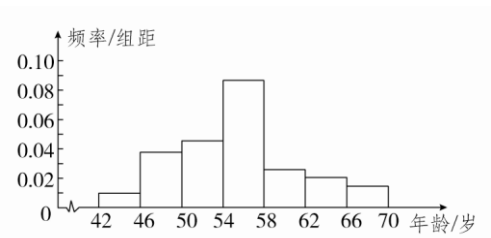
57,61,57,57,58,57,61,54,68,51,49,64,50,48,65,
52,56,46,54,49,51,47,55,55,54,42,51,56,55,51,
54,51,60,62,43,55,56,61,52,69,64,46,54,47,70.

- (1)将数据分为 7 组,列出频率分布表,并画出相应的频率分布直方图;
(2)用自己的语言描述一下历届美国总统就任时年龄的分布情况.

解 (1)以 4 为组距,列频率分布表如下:

分组	频数	频率
[42,46)	2	0.0444
[46,50)	7	0.1555
[50,54)	8	0.1778
[54,58)	16	0.3556
[58,62)	5	0.1111
[62,66)	4	0.0889
[66,70]	3	0.0667
合计	45	1.0000

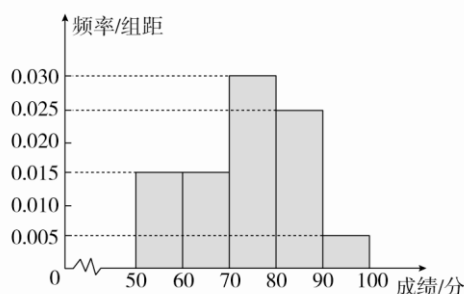
画出相应的频率分布直方图,如图所示.



(2)从频率分布表中可以看出,将近 60%的美国总统就任时的年龄在 50 岁至 60 岁之间,45 岁及 45 岁以下和 65 岁以上就任的总统所占的比例相对较小.

题型二 频率分布直方图的应用

例 2 为了迎接某市作为全国文明城市的复查,爱卫会随机抽取了 60 位路人进行问卷调查,调查项目是自己对该市各方面卫生情况的满意度(假设被问卷的路人回答是客观的),以分数表示问卷结果,并统计他们的问卷分数,把其中不低于 50 分的分成五段: $[50,60)$, $[60,70)$, $\cdots$, $[90,100]$ 后画出如图所示的部分频率分布直方图,观察图形信息,回答下列问题:



- (1) 求出问卷调查分数低于 50 分的被问卷人数;
- (2) 估计全市市民满意度在 60 分及以上的百分比.

[解] (1) 因为各组的频率之和等于 1,

故低于 50 分的频率为

$f = 1 - (0.015 \times 2 + 0.030 + 0.025 + 0.005) \times 10 = 0.1$, 故低于 50 分的人数为 $60 \times 0.1 = 6$.

(2) 依题意, 60 分及以上的频率和为

$$(0.015 + 0.030 + 0.025 + 0.005) \times 10 = 0.75,$$

所以抽样满意度在 60 分及以上的百分比为 75%.

于是, 可以估计全市市民满意度在 60 分及以上的百分比为 75%.

金版点睛 频率分布直方图的应用

频率分布指的是一个样本数据在各个小范围内所占比例的大小, 一般用频率分布直方图反映样本的频率分布, 其中:

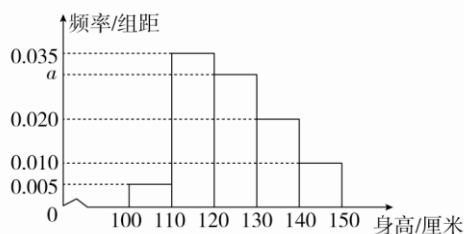
(1) 频率分布直方图中纵轴表示 $\frac{\text{频率}}{\text{组距}}$;

(2) 频率分布直方图中, 各个小长方形的面积等于频率, 各个小长方形的面积之和为 1;

(3) 长方形的高的比也就是频率之比.

跟踪训练2

从某小学随机抽取 100 名同学, 将他们的身高(单位: 厘米)数据绘制成频率分布直方图(如图所示), 由图中数据可知 $a = \underline{\hspace{2cm}}$. 若要从身高在 $[120, 130)$, $[130, 140)$, $[140, 150]$ 三组内的学生中, 用分层抽样的方法选取 18 人参加一项活动, 则从身高在 $[140, 150]$ 内的学生中选取的人数应为 $\underline{\hspace{2cm}}$.

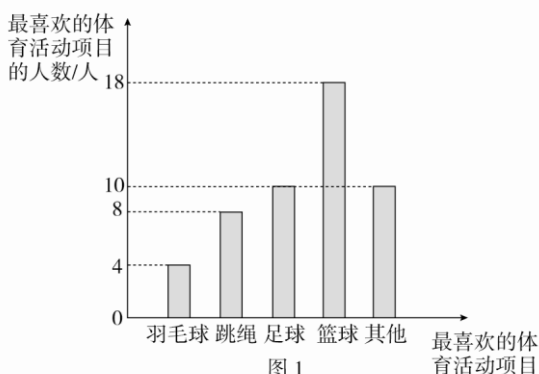


答案 0.030 3

解析 因为频率分布直方图中各小长方形的面积之和为 1, 所以 $10 \times (0.005 + 0.035 + a + 0.020 + 0.010) = 1$, 解得 $a = 0.030$. 由图可知身高在 $[120, 150]$ 内的学生人数为 $100 \times 10 \times (0.030 + 0.020 + 0.010) = 60$, 其中身高在 $[140, 150]$ 内的学生人数为 10, 所以从身高在 $[140, 150]$ 内的学生中选取的人数为 $\frac{18}{60} \times 10 = 3$.

题型三 统计图表的应用

例 3 某省有关部门要求各中小学要把“每天锻炼一小时”写入课程表, 为了响应这一号召, 某校围绕着“你最喜欢的体育活动项目是什么? (只写一项)”的问题, 对在校学生进行了随机抽样调查, 从而得到一组数据. 图 1 是根据这组数据绘制的柱形图. 请结合柱形图回答下列问题:



(1) 该校对多少名学生进行了抽样调查?

(2) 本次抽样调查中, 最喜欢篮球活动的有多少人? 占被调查人数的百分比是多少?

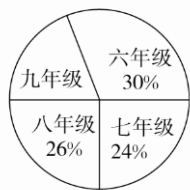


图 2

(3)若该校九年级共有 200 名学生，图 2 是根据各年级学生人数占全校学生总人数的百分比绘制的扇形图，请你估计全校学生中最喜欢跳绳活动的人数约为多少？

[解] (1)由图 1 知 $4 + 8 + 10 + 18 + 10 = 50$ (名)，所以该校对 50 名学生进行了抽样调查。

(2)本次调查中，最喜欢篮球活动的有 18 人，占被调查人数的 $\frac{18}{50} \times 100\% = 36\%$ 。

(3) $1 - (30\% + 26\% + 24\%) = 20\%$,

$200 \div 20\% = 1000$ (人)，

$\frac{8}{50} \times 100\% \times 1000 = 160$ (人)，

所以估计全校学生中最喜欢跳绳活动的人数约为 160。

金版点睛

(1)柱形图是用一个单位长度表示一定的数量，根据数量的多少画成高度不同的小矩形，然后把这些小矩形按照一定的顺序排列起来。其特点是便于看出和比较各种数量的多少，即柱形图能清楚地表示出每个项目的具体数目。

(2)扇形图是用整个圆面积表示总数(100%)，用圆内的扇形面积表示各部分所占总数的百分数。

总之，用图表来表示数量关系更生动形象、具体，使人一目了然。

跟踪训练3

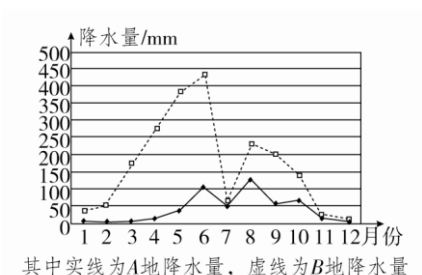
下表给出了 2018 年 A, B 两地的降水量(单位: mm):

	1 月	2 月	3 月	4 月	5 月	6 月
A	9.2	4.9	5.4	18.6	38.0	106.3
B	41.4	53.3	178.8	273.5	384.9	432.4
	7 月	8 月	9 月	10 月	11 月	12 月
A	54.4	128.9	62.9	73.6	26.2	10.6
B	67.5	228.5	201.4	147.3	28.0	19.1

(1)根据统计表绘制折线图；

(2)根据折线图比较 A, B 两地的降水量, 分析哪个地方的降水量较大?

解 (1)建立直角坐标系, 用横坐标上的点表示月份, 用纵坐标上的点表示降水量, 描出每个月份对应的点, 然后用直线段顺次连接相邻的点, 得到折线统计图如图表示.



(2)观察折线图, 从整体上看, B 地降水量较大.

随堂水平达标

1. 反映某种股票的涨跌情况, 应选择()

A. 条形图

B. 折线图

C. 扇形图

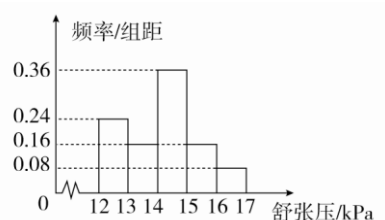
D. 三种图均可

答案 B

解析 条形图主要用于直观描述不同类别或分组数据的频数, 折线图主要用于描述数据随时间变化的趋势, 扇形图主要用于直观描述各类数据占总数的比例. 故选 B.

2. 为了研究某药品的疗效, 选取若干名志愿者进行临床试验, 所有志愿者的舒张压数据(单位: kPa)的分组区间为 $[12,13)$, $[13,14)$, $[14,15)$, $[15,16)$, $[16,17]$, 将其按从左到右的顺序分别编号为第一组, 第二组, $\dots$, 第五组, 如图是根据试

验数据制成的频率分布直方图. 已知第一组与第二组共有 20 人, 第三组中没有疗效的有 6 人, 则第三组中有疗效的人数为()



- A. 6
B. 8
C. 12
D. 18

答案 C

解析 志愿者的总人数为 $\frac{20}{(0.24 + 0.16) \times 1} = 50$ ，所以第三组的人数为

$50 \times 0.36 \times 1 = 18$, 所以有疗效的人数为 $18 - 6 = 12$.

3. 一个频率分布表(样本量为 30)不小心被损坏了一部分, 只记得样本中数据在 $[20, 60)$ 内的频率为 0.8, 则估计样本在 $[40, 50)$, $[50, 60)$ 内的数据个数共为()

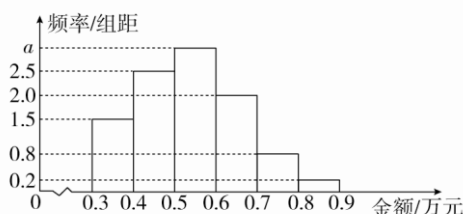
分组	[10,20)	[20,30)	[30,40)	
频数	3	4	5	

- A. 14
B. 15
C. 16
D. 17

答案 B

解析 ∵ 样本中数据在 $[20, 60)$ 内的频率为 0.8 , ∴ 样本数据在 $[20, 60)$ 内的频数为 $30 \times 0.8 = 24$, ∴ 样本在 $[40, 50)$, $[50, 60)$ 内的数据个数共为 $24 - 4 - 5 = 15$.

4. 某电子商务公司对 10000 名网络购物者 2019 年度的消费情况进行统计, 发现消费金额(单位: 万元)都在区间 $[0.3, 0.9]$ 内, 其频率分布直方图如图所示.



- (1)直方图中的 $a=$ _____;
- (2)在这些购物者中, 消费金额在区间 $[0.5,0.9]$ 内的购物者的人数为_____.

答案 (1)3.0 (2)6000

解析 由频率分布直方图及频率和等于 1 可得 $0.2 \times 0.1 + 0.8 \times 0.1 + 1.5 \times 0.1$

$+ 2.0 \times 0.1 + 2.5 \times 0.1 + a \times 0.1 = 1$, 解得 $a = 3.0$, 消费金额在区间 $[0.5, 0.9]$ 内的频率为 $0.2 \times 0.1 + 0.8 \times 0.1 + 2.0 \times 0.1 + 3.0 \times 0.1 = 0.6$, 所以消费金额在区间 $[0.5, 0.9]$ 内的购物者的人数为 $0.6 \times 10000 = 6000$.

5. 从高三参加数学竞赛的学生中抽取 50 名学生的成绩, 成绩的分组及各组的频数如下(单位: 分):

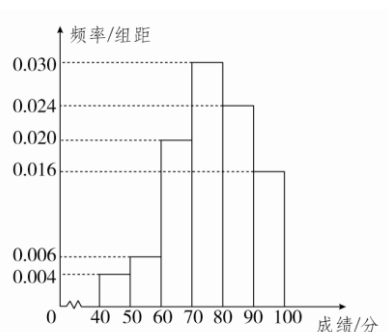
$[40, 50)$, 2; $[50, 60)$, 3; $[60, 70)$, 10; $[70, 80)$, 15; $[80, 90)$, 12; $[90, 100]$, 8.

- (1)列出样本的频率分布表;
- (2)画出频率分布直方图;
- (3)估计成绩在 $[60, 90)$ 分的学生比例;
- (4)估计成绩在 80 分以下的学生比例.

解 (1)频率分布表如下:

成绩分组	频数	频率
$[40, 50)$	2	0.04
$[50, 60)$	3	0.06
$[60, 70)$	10	0.2
$[70, 80)$	15	0.3
$[80, 90)$	12	0.24
$[90, 100]$	8	0.16
合计	50	1

(2)频率分布直方图如图所示.



(3)样本中成绩在 $[60, 90)$ 分的学生比例为 $0.20 + 0.30 + 0.24 = 0.74 = 74\%$.

由样本估计总体, 成绩在 $[60, 90)$ 分的学生约占 74%.

(4)样本中成绩在 80 分以下学生比例为 $1 - (0.24 + 0.16) = 1 - 0.4 = 0.6 = 60\%$.

由样本估计总体, 成绩在 80 分以下的学生约占 60%.



9.2.4 总体离散程度的估计

新课标新学法 (教师独具内容)

课程标准: 1. 结合实例,能用样本估计总体的离散程度参数(标准差、方差、极差). 2. 理解离散程度参数的统计含义.

教学重点: 理解样本数据方差、标准差的意义和作用并会计算方差、标准差.

教学难点: 从样本数据中提取基本的数字特征,并作出合理解释,能估计总体的离散程度.

核心概念掌握



知识导学

知识点一 一组数据的方差与标准差

假设一组数据是 $x_1, x_2, \dots, x_n$, 用 $\bar{x}$ 表示这组数据的平均数, 我们称 $\textcircled{01} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 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 $\textcircled{02} \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$ 的形式.

方差的算术平方根 $\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$ 称为这组数据的 $\textcircled{03}$ 标准差.

知识点二 总体方差与总体标准差

如果总体中所有个体的变量值分别为 $Y_1, Y_2, \dots, Y_N$, 总体平均数为 $\bar{Y}$, 则称 $\boxed{01} S^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$ 为总体方差, $\boxed{02} S = \sqrt{S^2}$ 为总体标准差. 与总体均值类似, 总体方差也可以写成加权的形式. 如果总体的 N 个变量值中, 不同的值共有 $k (k \leq N)$ 个, 不妨记为 $Y_1, Y_2, \dots, Y_k$, 其中 Y_i 出现的频数为 $f_i (i=1, 2, \dots, k)$, 则总体方差为 $\boxed{03} S^2 = \frac{1}{N} \sum_{i=1}^k f_i (Y_i - \bar{Y})^2$.

知识点三 样本方差与样本标准差

如果一个样本中个体的变量值分别为 $y_1, y_2, \dots, y_n$, 样本平均数为 $\bar{y}$, 则称 $\boxed{01} s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ 为样本方差, $\boxed{02} s = \sqrt{s^2}$ 为样本标准差.

知识点四 标准差、方差描述数据的特征

标准差刻画了数据的 $\boxed{01}$ 离散程度或 $\boxed{02}$ 波动幅度, 标准差越大, 数据的离散程度 $\boxed{03}$ 越大; 标准差越小, 数据的离散程度 $\boxed{04}$ 越小. 在刻画数据的分散程度上, 方差和标准差是一样的. 但在解决实际问题中, 一般多采用标准差.

知识点五 分层随机抽样估计总体方差

设层数为 2 层的分层随机抽样, 第 1 层和第 2 层包含的样本变量由 $x_1, x_2, \dots, x_n$ 及 $y_1, y_2, \dots, y_n$ 表示.

	样本数	总体数	方差	平均数
第 1 层	m	M	s_x^2	$\bar{x}$
第 2 层	n	N	s_y^2	$\bar{y}$

则总体方差 $s^2 = \frac{M[s_x^2 + (\bar{x} - \bar{z})^2] + N[s_y^2 + (\bar{y} - \bar{z})^2]}{M + N}$

注： $\bar{x}$ 为总体平均数，即 $\bar{x} = \frac{\sum_{i=1}^M x_i + \sum_{i=1}^N y_i}{M+N}$.



新知拓展

1. 方差的简化计算公式： $s^2 = \frac{1}{n}[(x_1^2 + x_2^2 + \cdots + x_n^2) - n\bar{x}^2]$ ，或写成 $s^2 = \frac{1}{n}(x_1^2 + x_2^2 + \cdots + x_n^2) - \bar{x}^2$. 即方差等于原数据平方的平均数减去平均数的平方.

2. 平均数、方差公式的推广

(1) 若数据 $x_1, x_2, \cdots, x_n$ 的平均数为 $\bar{x}$ ，那么 $mx_1 + a, mx_2 + a, \cdots, mx_n + a$ 的平均数是 $m\bar{x} + a$.

(2) 若数据 $x_1, x_2, \cdots, x_n$ 的方差为 s^2 ，那么

① 数据 $x_1 + a, x_2 + a, \cdots, x_n + a$ 的方差也是 s^2 ;

② 数据 $ax_1, ax_2, \cdots, ax_n$ 的方差是 a^2s^2 .

☒ 评价自测

1. 判一判(正确的打“√”，错误的打“×”)

(1) 方差越大，数据的稳定性越强. ()

(2) 在两组数据中，平均值较大的一组方差较大. ()

(3) 样本的平均数和标准差一起反映总体数据的取值信息. 一般地，绝大部分数据落在 $[\bar{x} - 2s, \bar{x} + 2s]$ 内. ()

(4) 平均数反映数据的集中趋势，方差则反映数据离平均值的波动大小. ()

答案 (1)× (2)× (3)√ (4)√

2. 做一做

(1) 下列说法不正确的是()

A. 方差是标准差的平方

B. 标准差的大小不会超过极差

C. 若一组数据的值大小相等，没有波动变化，则标准差为 0

D. 标准差越大，表明各个样本数据在样本平均数周围越集中；标准差越小，表明各个样本数据在样本平均数周围越分散

(2) 某学员在一次射击测试中射靶 10 次，命中环数如下：

7, 8, 7, 9, 5, 4, 9, 10, 7, 4.

则：① 平均命中环数为_____；

② 命中环数的标准差为_____.

(3)样本中共有五个个体,其值分别为 $a, 0, 1, 2, 3$, 若该样本的平均值为 1, 则该样本的方差为_____.

答案 (1)D (2)①7 ②2 (3)2

核心素养形成

题型一 样本的标准差与方差的求法

例 1 从甲、乙两种玉米中各抽 10 株, 分别测得它们的株高如下:

甲: 25, 41, 40, 37, 22, 14, 19, 39, 21, 42;

乙: 27, 16, 44, 27, 44, 16, 40, 40, 16, 40;

试计算甲、乙两组数据的方差和标准差.

[解] $\bar{x}_{\text{甲}} = \frac{1}{10} \times (25 + 41 + 40 + 37 + 22 + 14 + 19 + 39 + 21 + 42) = 30,$

$$s_{\text{甲}}^2 = \frac{1}{10} \times [(25 - 30)^2 + (41 - 30)^2 + (40 - 30)^2 + (37 - 30)^2 + (22 - 30)^2 + (14 - 30)^2 + (19 - 30)^2 + (39 - 30)^2 + (21 - 30)^2 + (42 - 30)^2] = 104.2,$$

$$s_{\text{甲}} = \sqrt{104.2} \approx 10.208.$$

$$\bar{x}_{\text{乙}} = \frac{1}{10} \times (27 + 16 + 44 + 27 + 44 + 16 + 40 + 40 + 16 + 40) = 31,$$

$$\text{同理 } s_{\text{乙}}^2 = 128.8,$$

$$s_{\text{乙}} = \sqrt{128.8} \approx 11.349.$$

金版点睛 对标准差与方差概念的理解

(1)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.

(2)标准差、方差的取值范围: $[0, +\infty)$.

标准差、方差为 0 时, 样本各数据全相等, 表明数据没有波动幅度, 数据没有离散性.

(3)因为方差与原始数据的单位不同, 且平方后可能放大了偏差的程度, 所以虽然方差与标准差在刻画样本数据的离散程度上是一样的, 但在解决实际问题时, 一般多采用标准差.

跟踪训练1

某班 40 名学生平均分成两组, 两组学生某次考试成绩情况如下表所示:

组别	平均数	标准差
第一组	90	4
第二组	80	6

求这次考试成绩的平均数和标准差.

$$\left(\begin{array}{l} \text{注: 标准差 } s = \sqrt{\frac{1}{n}[(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2]} \\ = \sqrt{\frac{1}{n}[(x_1^2 + x_2^2 + \cdots + x_n^2) - n\bar{x}^2]} \end{array} \right)$$

解 设第一组数据为 $x_1, x_2, \cdots, x_{20}$, 第二组数据为 $x_{21}, x_{22}, \cdots, x_{40}$, 全班平均成绩为 $\bar{x}$.

$$\text{根据题意, 有 } \bar{x} = \frac{90 \times 20 + 80 \times 20}{40} = 85,$$

$$4^2 = \frac{1}{20}(x_1^2 + x_2^2 + \cdots + x_{20}^2 - 20 \times 90^2),$$

$$6^2 = \frac{1}{20}(x_{21}^2 + x_{22}^2 + \cdots + x_{40}^2 - 20 \times 80^2),$$

$$\therefore x_1^2 + x_2^2 + \cdots + x_{40}^2 = 20 \times (4^2 + 6^2 + 90^2 + 80^2) = 291040.$$

$$\text{再由变形公式, 得 } s^2 = \frac{1}{40}(x_1^2 + x_2^2 + \cdots + x_{40}^2 - 40\bar{x}^2)$$

$$= \frac{1}{40}(x_1^2 + x_2^2 + \cdots + x_{40}^2 - 40 \times 85^2) = \frac{1}{40} \times (291040 - 289000) = 51,$$

$$\therefore s = \sqrt{51}.$$

题型二 样本标准差、方差的实际应用

例 2 某工厂甲、乙两名工人参加操作技能培训, 他们在培训期间参加的 8 次测试成绩记录如下:

甲: 95 82 88 81 93 79 84 78

乙: 83 92 80 95 90 80 85 75

(1) 试比较哪个工人的成绩较好;

(2) 甲、乙成绩位于 $\bar{x} - s$ 与 $\bar{x} + s$ 之间有多少?

$$\text{[解]} \quad (1) \quad \bar{x}_{\text{甲}} = \frac{1}{8} \times (95 + 82 + 88 + 81 + 93 + 79 + 84 + 78) = 85,$$

$$\bar{x}_{\text{乙}} = \frac{1}{8} \times (83 + 92 + 80 + 95 + 90 + 80 + 85 + 75) = 85.$$

$$s_{\text{甲}}^2 = \frac{1}{8} \times [(95 - 85)^2 + (82 - 85)^2 + (88 - 85)^2 + (81 - 85)^2 + (93 - 85)^2 + (79 - 85)^2 + (84 - 85)^2 + (78 - 85)^2] = 35.5,$$

$$s_{\text{乙}}^2 = \frac{1}{8} \times [(83 - 85)^2 + (92 - 85)^2 + (80 - 85)^2 + (95 - 85)^2 + (90 - 85)^2 + (80 - 85)^2 + (85 - 85)^2 + (75 - 85)^2] = 41.$$

$$\because \bar{x}_{\text{甲}} = \bar{x}_{\text{乙}}, s_{\text{甲}}^2 < s_{\text{乙}}^2,$$

$\therefore$ 甲的成绩较稳定.

综上可知, 甲的成绩较好.

$$(2) \because s_{\text{甲}} = \sqrt{s_{\text{甲}}^2} = \sqrt{35.5} \approx 5.96,$$

$$\bar{x}_{\text{甲}} - s_{\text{甲}} = 79.04, \quad \bar{x}_{\text{甲}} + s_{\text{甲}} = 90.96,$$

$\therefore$ 甲位于 $[\bar{x} - s, \bar{x} + s]$ 之间的数据有 4 个.

$$\text{又 } s_{\text{乙}} = \sqrt{s_{\text{乙}}^2} = \sqrt{41} \approx 6.4,$$

$$\bar{x}_{\text{乙}} - s_{\text{乙}} = 78.6, \quad \bar{x}_{\text{乙}} + s_{\text{乙}} = 91.4,$$

$\therefore$ 乙的成绩位于 $[\bar{x} - s, \bar{x} + s]$ 之间的有 5 个.

金版点睛

比较两组数据的异同点, 一般情况是从平均数及标准差这两个方面考虑. 其中标准差与样本数据单位一样, 比方差更直观地刻画出与平均数的平均距离.

跟踪训练2

从甲、乙两名学生中选拔一人参加射箭比赛, 为此需要对他们的射箭水平进行测试. 现这两名学生在相同条件下各射箭 10 次, 命中的环数如下:

甲	8	9	7	9	7	6	10	10	8	6
乙	10	9	8	6	8	7	9	7	8	8

(1) 计算甲、乙两人射箭命中环数的平均数和标准差;

(2) 比较两个人的成绩, 然后决定选择哪名学生参加射箭比赛.

解 (1) 根据题中所给数据, 可得甲的平均数为

$$\bar{x}_{\text{甲}} = \frac{1}{10} \times (8 + 9 + 7 + 9 + 7 + 6 + 10 + 10 + 8 + 6) = 8,$$

乙的平均数为 $\bar{x}_Z = \frac{1}{10} \times (10 + 9 + 8 + 6 + 8 + 7 + 9 + 7 + 8 + 8) = 8$,

甲的标准差为

$$s_{\text{甲}} = \sqrt{\frac{1}{10} \times [(8-8)^2 + (9-8)^2 + \cdots + (6-8)^2]} = \sqrt{2},$$

乙的标准差为

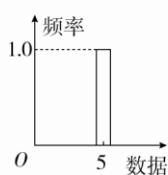
$$s_Z = \sqrt{\frac{1}{10} \times [(10-8)^2 + (9-8)^2 + \cdots + (8-8)^2]} = \frac{\sqrt{30}}{5},$$

故甲的平均数为 8, 标准差为 $\sqrt{2}$, 乙的平均数为 8, 标准差为 $\frac{\sqrt{30}}{5}$.

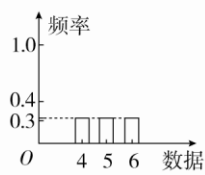
(2) $\because \bar{x}_{\text{甲}} = \bar{x}_Z$, 且 $s_{\text{甲}} > s_Z$, $\therefore$ 乙的成绩较为稳定, 故选择乙参加射箭比赛.

题型三 标准差、方差的图形分析

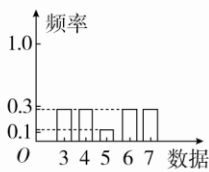
例 3 样本数为 9 的四组数据, 他们的平均数都是 5, 条形图如下图所示, 则标准差最大的一组是()



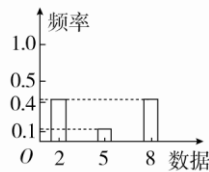
第一组



第二组



第三组



第四组

A. 第一组

B. 第二组

C. 第三组

D. 第四组

【解析】 第一组中, 样本数据都为 5, 数据没有波动幅度, 标准差为 0; 第二组中, 样本数据为 4, 4, 4, 5, 5, 5, 6, 6, 6, 标准差为 $\frac{\sqrt{6}}{3}$; 第三组中, 样本数据为 3, 3, 4, 4, 5, 6, 6, 7, 7, 标准差为 $\frac{2\sqrt{5}}{3}$; 第四组中, 样本数据为 2, 2, 2, 2, 5, 8, 8, 8, 8, 标准差为 $2\sqrt{2}$, 故标准差最大的一组是第四组.

【答案】 D

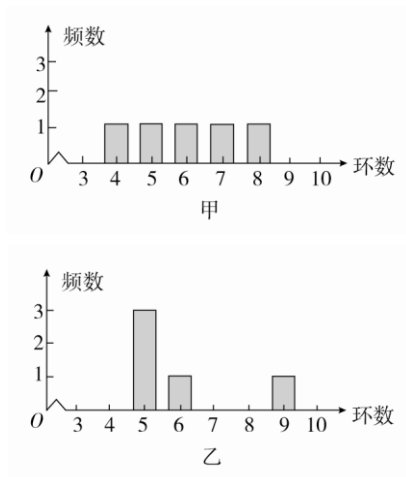
金版点睛 由图形分析标准差、方差的大小

从四个图形可以直观看出第一组数据没有波动性, 第二、三组数据的波动性

都比较小，而第四组数据的波动性相对较大，利用标准差的意义也可以直观得到答案.

跟踪训练3

甲、乙两人在一次射击比赛中各射靶 5 次，两人成绩的条形统计图如图所示，则()



- A. 甲的成绩的平均数小于乙的成绩的平均数
- B. 甲的成绩的中位数等于乙的成绩的中位数
- C. 甲的成绩的方差小于乙的成绩的方差
- D. 甲的成绩的极差小于乙的成绩的极差

答案 C

解析 由题图可得，甲的成绩为 4,5,6,7,8，乙的成绩为 5,5,5,6,9，所以甲、乙的成绩的平均数均是 6，故 A 不正确；甲、乙成绩的中位数分别为 6,5，故 B 不正确；甲、乙成绩的极差都是 4，故 D 不正确；甲的成绩的方差为 $\frac{1}{5} \times (2^2 \times 2 + 1^2 \times 2) = 2$ ，乙的成绩的方差为 $\frac{1}{5} \times (1^2 \times 3 + 3^2) = 2.4$. 故 C 正确.

随堂水平达标

- 下列选项中，能反映一组数据的离散程度的是()
 - A. 平均数
 - B. 中位数
 - C. 方差
 - D. 众数

答案 C

解析 由方差的定义，知方差反映了一组数据的离散程度.

- 样本数据 2,4,6,8,10 的标准差为()
 - A. 40
 - B. 8

C. $2\sqrt{10}$

D. $2\sqrt{2}$

答案 D

解析 $\bar{x} = \frac{1}{5} \times (2 + 4 + 6 + 8 + 10) = 6$, 则标准差为

$$\sqrt{\frac{1}{5} \times [(2 - 6)^2 + (4 - 6)^2 + (6 - 6)^2 + (8 - 6)^2 + (10 - 6)^2]} = 2\sqrt{2}.$$

3. 甲、乙、丙、丁四人参加某运动会射击项目选拔赛, 四人的平均成绩和方差如下表所示:

	甲	乙	丙	丁
平均环数 $\bar{x}$	8.3	8.8	8.8	8.7
方差 s^2	3.5	3.6	2.2	5.4

若要从这四人中选择一人去参加该运动会射击项目比赛, 最佳人选是_____. (填“甲”“乙”“丙”“丁”中的一个)

答案 丙

解析 分析题中表格数据可知, 乙与丙的平均环数最多, 又丙的方差比乙小, 说明丙成绩发挥得较为稳定, 所以最佳人选为丙.

4. 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_____.

答案 0.1

解析 这组数据的平均数

$$\bar{x} = \frac{4.7 + 4.8 + 5.1 + 5.4 + 5.5}{5} = 5.1,$$

则方差 $s^2 =$

$$\frac{(4.7 - 5.1)^2 + (4.8 - 5.1)^2 + (5.1 - 5.1)^2 + (5.4 - 5.1)^2 + (5.5 - 5.1)^2}{5}$$

$$= \frac{0.16 + 0.09 + 0 + 0.09 + 0.16}{5} = 0.1.$$

5. 甲、乙两机床同时加工直径为 100 cm 的零件, 为检验质量, 各从中抽取 6 件测量, 数据为:

甲: 99 100 98 100 100 103

乙: 99 100 102 99 100 100

(1)分别计算两组数据的平均数及方差；

(2)根据计算结果判断哪台机床加工零件的质量更稳定.

解 (1) $\bar{x}_{\text{甲}} = \frac{1}{6} \times (99 + 100 + 98 + 100 + 100 + 103) = 100,$

$$\bar{x}_{\text{乙}} = \frac{1}{6} \times (99 + 100 + 102 + 99 + 100 + 100) = 100.$$

$$s_{\text{甲}}^2 = \frac{1}{6} \times [(99 - 100)^2 + (100 - 100)^2 + (98 - 100)^2 + (100 - 100)^2 + (100 - 100)^2 +$$

$$(103 - 100)^2] = \frac{7}{3},$$

$$s_{\text{乙}}^2 = \frac{1}{6} \times [(99 - 100)^2 + (100 - 100)^2 + (102 - 100)^2 + (99 - 100)^2 + (100 - 100)^2 +$$

$$(100 - 100)^2] = 1.$$

(2)两台机床所加工零件的直径的平均数相同,

又 $s_{\text{甲}}^2 > s_{\text{乙}}^2$, 所以乙机床加工零件的质量更稳定.