

第六章 统计

S1 获取数据的途径 & S2 抽样的基本方法

基础满分

1. C 【解析】调查一个地区糖尿病的发病率,调查数量较多,不适合全面调查,故 A 错误;

了解一批水稻种子的发芽率,调查数目较多,且具有破坏性,不适合全面调查,故 B 错误;

了解一个班级学生的身高情况,数量较少,操作方便,适合全面调查,故 C 正确;

了解某城市居民的生活水平,调查数目较多,不适合全面调查,故 D 错误.

2. ACD 【解析】若样本数量较小,则适宜全面调查,故 B 错误;若调查范围较广,经济成本较高,或调查具有破坏性,则不适宜全面调查,适宜抽样调查,故 ACD 正确.

3. A 【解析】由题可知,为了分析该校高一年级 1 800 名学生的学习成绩,从中随机抽取了 180 名学生的成绩单,则样本容量为 180,故 A 正确;

每名学生的成绩是所抽取的一个个体,故 B, C 错误;

1 800 名学生的成绩是总体,故 D 错误.

4. BD 【解析】简单随机抽样中被抽取样本的总体的个数有限,它是从总体中逐个随机抽取,样本抽取是不放回抽样,简单随机抽样是一种等可能抽样,即每个个体被抽取的可能性相等. 故 BD 错误,故选 BD.

5. B 【解析】总体的个数较大,不适合用抽签法,故 A 错误;

总体容量和样本容量都较小,且同厂生产的产品可视为搅拌均匀,故 B 正确;

甲、乙两厂生产的产品质量可能差别较大,因此未达到搅拌均匀的条件,故 C 错误;

总体的个数较大,不适合用抽签法,故 D 错误.

6. C 【解析】被选中的红色球号码依次为 28, 03, 22, 24, ..., 所以第四个被选中的红色球号码为 24. 故 C 正确.

7. AD 【解析】从 50 个零件中逐个抽取 5 个做质量检验,符合简单随机抽样要求,故 A 正确;

老师表扬的是发言积极的,对每一个个体而言,不具备“等可能性”,故 B 错误;

总体容量是无限的,不符合简单随机抽样要求,故 C 错误;

8 条跑道,抽取 1 条,总体有限,每个个体被抽到的机会均等,是简单随机抽样,故 D 正确.

易错警示 忽略简单随机抽样的特点

满足总体有限性、样本有限性、不放回抽样、等可能性四个特点的抽样才是简单随机抽样,缺一不可.

8. 【解】(1) 将 50 名志愿者编号,号码分别是 1, 2, ..., 50;

(2) 将号码分别写在外观、质地等无差别的小纸片上作为号签;

(3) 将小纸片放入一个不透明的盒子里,充分搅匀;

(4) 从盒子中不放回地逐个抽取 8 个号签,使与号签上编号对应的志愿者进入样本,组成志愿服务小组.

9. 【解】抽签法:

第一步:对 450 户居民编号(可使用门牌号);

第二步:准备 450 个完全相同的号签分别标上这些编号,将号签放在一个不透明的容器中,搅拌均匀,每次从中抽取一个号签,不放回地抽取 50 次;

第三步:将取出的 50 个号签上的号码所对应的 50 户居民的月均用水量组成样本.

随机数法:

第一步:对 450 户居民编号,如按 001~450 进行编号;

第二步:用随机数工具产生 001~450 范围内的整数值随机数;

第三步:把产生的随机数作为抽中的编号,把与编号对应的居民的月均用水量录入样本,直到抽出 50 户居民的月均用水量.

两种抽样方法对比:抽签法简单易行,但当总体较大时,操作起来比较麻烦,很难使总体“搅拌均匀”,所以抽签法适用于总体个数不多的情形;

随机数法适用于总体较多的情形,主要优点是节省人力、物力、财力和时间,缺点是当总体很大时,用随机数法给总体编号等准备工作很费事,抽中的个体可能会比较分散等.

10. C 【解析】常用的抽样方法有简单随机抽样和分层随机抽样,由题可知该地区小学、初中、高中三个学段学生的肺活量有较大差异,而同一学段男、女生肺活量差异不大,所以最合理的抽样方法是按学段分层随机抽样. 故 C

正确.

- 11. A** 【解析】职工共有 500 人,其中男性职工有 320 人,采用分层抽样的方法抽取 100 人,则抽取到的男性职工人数为 $320 \times \frac{100}{500} = 64$, 女性职工人数为 $100 - 64 = 36$, 则抽取到的男性职工的人数比女性职工的人数多 $64 - 36 = 28$. 故 A 正确.

- 12. A** 【解析】张三与李四被抽到的可能性一样大,故 A 错误;
理学专业应抽取的人数为 $100 \times \frac{30}{100} = 30$, 工学专业应抽取的人数为 $100 \times \frac{20}{100} = 20$, 故 B 正确;

由题图知,各专业之间差异比较大,所以采用分层随机抽样更合理,故 C 正确;

该问题中的样本容量为 100, 故 D 正确.

- 13. 300** 【解析】抽样比为 $\frac{80}{480} = \frac{1}{6}$, 样本中女生的人数为 $80 - 30 = 50$, 所以该校女生的人数为 $\frac{50}{\frac{1}{6}} = 300$.

- 14. 【解】**全校参与跳绳的人数占总人数的 $\frac{2}{5}$, 则跳绳的人数为 $\frac{2}{5} \times 2\,000 = 800$, 所以跑步的人数为 $\frac{3}{5} \times 2\,000 = 1\,200$. 又因为 $a : b : c = 2 : 3 : 5$, 所以 $a = \frac{2}{10} \times 1\,200 = 240$, $b = \frac{3}{10} \times 1\,200 = 360$, $c = \frac{5}{10} \times 1\,200 = 600$. 因为抽取样本为 200 人, 即抽样比例为

$\frac{200}{2\,000} = \frac{1}{10}$, 所以高二年级中参与跑步的同学应抽取 $360 \times \frac{1}{10} = 36$ (人).

- 15. B** 【解析】由于总体容量相对较大, 样本量较少, 故采用随机数法较为合适. 故 B 正确.

- 16. C** 【解析】因为社会购买力的某项指标, 受到家庭收入的影响, 而社区中家庭收入差别明显, 所以要从中抽取一个容量为 100 的样本, 应采用分层随机抽样. 故 C 正确.

- 17. C** 【解析】①中代理商的商店规模不同, 分层特征明显, 故应采用分层随机抽样;

②中总体差异不明显, 且总体中的个体数较少, 样本量较少, 故应采用简单随机抽样. 故 C 正确.

- 18. A** 【解析】①从 12 瓶饮料中抽取 4 瓶进行食品卫生检查, 总体容量较小, 样本容量较小, 宜用抽签法;

②某生活小区共有 540 名居民, 其中年龄不超过 30 岁的有 180 人, 年龄在超过 30 岁不超过 60 岁的有 270 人, 60 岁以上的有 90 人, 具有明显的分层特点, 宜用分层随机抽样. 故 A 正确.

- 19. D** 【解析】若用分层随机抽样, 则在 001 到 108 编号中有 4 个, 109 到 189 中有 3 个, 190 到 270 中有 3 个, 可知①②③都有可能为分层随机抽样, ④不可能为分层随机抽样;

而抽签法和随机数法抽出的样本编号没有明显的特征, 故①②③④都有可能是抽签法或随机数法, 故 D 正确.

- 20. D** 【解析】在每个饲养房各抽取 6 只, 由于各饲养房所养数量不一, 反而造成了各个个体入选可能性的不均衡, 故 A 错误;
保证了各个个体入选可能性相等, 但没有注意到处在四个不同环境中会产生差异, 不如采用分层随机抽样可靠性高, 且统一编号统一选择加大了工作量, 故 B 错误;
总体采用了分层随机抽样, 但在每层中没有考虑到个体的差异 (如健壮程度, 灵活程度), 貌似随机, 实则各个个体被抽到的可能性不同, 故 C 错误, D 正确.

§1 & §2 考点训练

- 1. D** 【解析】对于 A, C, 了解某市高一年级学生的身高情况和一批待售袋装牛奶的细菌数是否达标, 总量太大不适合普查, 故错误;
对于 B, 考虑到安全原因, 应该进行普查, 故错误;
对于 D, 根据抽样调查和普查的特点可知, 应选择抽样调查, 故正确.

- 2. A** 【解析】在统计里, 我们把所要考察对象的全体叫总体, 它包含数量、属性两个方面, 所以 1 400 名学生的心理健康情况是总体, 其中每个学生的心理健康情况是一个个体, 400 名学生的心理健康情况是一个样本, 400 为样本容量, 故 A 正确, BCD 错误.

- 3. B** 【解析】抽签法第一步要进行编号, 然后做号签, 放入容器, 最后逐个不放回地抽取号签, 所以这些步骤的先后顺序应为④①③②. 故 B 正确.

- 4. A** 【解析】由题意可得 808 (不合适), 436, 789 (不合适), 535, 577, 348, 994 (不合适), 837 (不合适),

522, 535 (重复), 578 为满足条件的第 6 个样本编号, 故 A 正确.

5. B 【解析】由题意可知, 要用分层随机抽样的方法从中抽取一个容量为 120 的样本, 则应抽取的高一年级学生的人数为 $120 \times \frac{3}{3+4+5} = 30$. 故 B 正确.

6. C 【解析】高二和高三年级获奖同学共 $120 - 120 \times \frac{6}{24} = 90$, 故 A 错误;
不能确定获得银奖和获得铜奖的人数, 故 B 错误;
高一年级获奖人数为 30, 获得金奖人数为 $120 \times \frac{4}{24} = 20$, 故获得金奖的同学可能都在高一年级, 故 C 正确;
获得金奖的人数为 20, 获得银奖和铜奖的人数共 100, 故获奖同学中金奖所占比例不可能最高, 故 D 错误.

7. D 【解析】方案①: C 校参加调研考试的学生中有 30 名数学培优生, 从这 30 名培优生中抽取 10 份试卷进行分析, 数量比较小, 适用于简单随机抽样;
方案②: 从 600 名学生的试卷中抽取一个容量为 200 的样本进行分析, 数量较大, 而且四所学校学生的水平各不相同, 按照比例抽取学生人数比较合适, 适用于分层随机抽样. 故 D 正确.

8. 【解】(1) 设该厂本月生产轿车 n 辆, 由题意得 $\frac{50}{n} = \frac{10}{100+300}$, $n = 2\,000$, 则 $z = 2\,000 - 100 - 300 - 150 - 450 - 600 = 400$.
(2) 设所抽取的样本中有 m 辆舒适型轿车, 则 $\frac{400}{1\,000} = \frac{m}{5}$, 解得 $m = 2$, 故在 C 类轿车中抽取 2 辆舒适

型轿车, 3 辆标准型轿车.

S3 用样本估计总体的分布

基础满分

1. C 【解析】由题表可知, 样本中质量大于 130 克的苹果数约为 $20 - 1 - 3 - 4 = 12$, 则估计这堆苹果中质量大于 130 克的苹果数约占苹果总数的 $\frac{12}{20} = 60\%$. 故 C 正确.

2. A 【解析】 \because 样本数据在 $[20, 60)$ 上的频率为 0.8, \therefore 样本数据在 $[20, 60)$ 上的频数是 $30 \times 0.8 = 24$, \therefore 估计样本在 $[40, 50)$, $[50, 60)$ 内的数据个数共为 $24 - 4 - 5 = 15$. 故 A 正确.

3. 【解】(1) 由题意可知, 第 2 组的频数为 $0.35 \times 100 = 35$, 第 3 组的频率为 $\frac{30}{100} = 0.3$.

(2) \because 第 3, 4, 5 组共有 60 名学生, \therefore 利用分层随机抽样的方法在 60 名学生中抽取 6 名学生, 则抽取的第 3 组人数为 $\frac{30}{60} \times 6 = 3$, 第 4 组人数为 $\frac{20}{60} \times 6 = 2$, 第 5 组人数为 $\frac{10}{60} \times 6 = 1$, \therefore 第 3, 4, 5 组每组分别抽取 3 人、2 人、1 人.

4. D 【解析】由频率分布直方图可得, 阅读时间是 8~10 小时的组频率为 $0.125 \times 2 = 0.25$, 则阅读时间是 8~10 小时的组频数为 $0.25 \times 120 = 30$, 即阅读时间是 8~10 小时的组频数和组频率分别是 30 和 0.25, 故 D 正确.

5. C 【解析】每周的课外活动时间不少于 14 h 的频率为 $2 \times (0.16 + 0.12 + 0.08) = 0.72$, 故所求人数 $N = 0.72 \times 200 = 144$. 故 C 正确.

易错警示 混淆频率分布直方

图纵轴的含义

在频率分布直方图中, 每一个小矩形的底边长都相等, 即等于组距, 高是 $\frac{\text{频率}}{\text{组距}}$, 而不是频率, $\frac{\text{频率}}{\text{组距}} \times \text{组距} = \text{频率}$.

6. 40 【解析】设中间一个小长方形的面积为 x , 则其他 8 个小长方形的面积的和为 $\frac{5}{2}x$. 根据频率分布直方图各小矩形面积之和为 1, 得 $x + \frac{5}{2}x = 1$, 则 $x = \frac{2}{7}$, 即中间一组的频率为 $\frac{2}{7}$, 所以中间一组的频数为 $140 \times \frac{2}{7} = 40$.

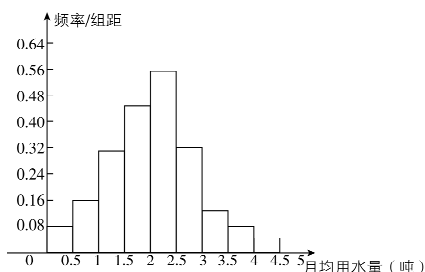
7. 【解】(1) 由频率分布直方图得, $(0.001 + 0.005 + 0.011 + 0.020 + 0.028 + 0.020 + 0.011 + a) \times 10 = 1$, 解得 $a = 0.004$.

(2) 因为样本中分数不低于 70 分的人数为 $(0.020 + 0.011 + 0.004) \times 10 \times 100 = 35$, 又因为样本中分数不低于 70 分的男生与女生人数之比为 4:3, 所以样本中分数不低于 70 分的男生人数为 $35 \times \frac{4}{7} = 20$, 因为样本中分数不低于 70 分的男生人数占样本中全部男生人数的 $\frac{1}{3}$, 所以样本中的男生人数为 $20 \times 3 = 60$, 故样本中女生人数为 $100 - 60 = 40$, 故样本中男女生人数之比为 3:2, 故总体中男生人数为 $1\,000 \times \frac{3}{5} = 600$, 女生人数为 $1\,000 - 600 = 400$.

8. 【解】(1) 在区间 $[0.5, 1)$ 内的频率为 0.08, 则频数为 8, 在区间 $[4, 4.5]$ 内的频率为 0.02, 则频数为 2, 则 $x = 100 - (4 + 8 + 15 + 22 + 14 +$

$$6+4+2=25, y=\frac{6}{100}=0.06.$$

(2) 绘制频率分布直方图如下,



因为从左往右数第4个矩形对应的频率为0.22,且表中的数据组距为0.5,所以所求高度为 $0.22 \div 0.5 = 0.44$.

- 9. ABD** 【解析】根据题意,由频率分布直方图可得到频率折线图,图示的 $\frac{\text{频率}}{\text{组距}}$ 表示的是各段分数的人数所占频率与组距的比值,而不是某个分数的人数所占频率与组距的比值,故A,B错误;
没有75到85分数段,故D错误;
70~90分的频率为 $\left(\frac{2}{55} + \frac{3}{110}\right) \times 10 = \frac{7}{11}$,所以分数在70~90分的人数为 $55 \times \frac{7}{11} = 35$,故C正确.

- 10. D** 【解析】共有2000户,需要抽取100户,故抽取的比例为 $\frac{100}{2000} = \frac{1}{20}$,由题中统计图可知,衣帽类有500户,果蔬类有300户,则衣帽类抽取 $500 \times \frac{1}{20} = 25$ 户,果蔬类抽取 $300 \times \frac{1}{20} = 15$ 户.故D正确.

- 11. C** 【解析】由题意可知,梧桐树苗有 $40\% \times 600 = 240$ 棵,根据人数占比可得中年教师应分得梧桐的数量为 $240 \times \frac{3}{5+3+2} = 72$ 棵.故C正确.

- 12. B** 【解析】由题意得,样本容量

为 $(350+450+200) \times 10\% = 100$,抽取C村贫困户的户数为 $200 \times 10\% \times 50\% = 10$.故B正确.

- 13. C** 【解析】根据物理的频率统计图,可知物理获二等奖的考生的频率为 $1 - 0.1 - 0.26 - 0.4 = 0.24$,又因为物理科目成绩为二等奖的考生有12人,所以该考场的人数为 $\frac{12}{0.24} = 50$,又根据化学的频率统计图,可知该考场化学考试获得一等奖的频率为 $1 - 0.16 - 0.38 - 0.38 = 0.08$,所以该考场化学考试获得一等奖的有 $50 \times 0.08 = 4$ 人,故①正确;
该考场物理考试获得二等奖的频率为0.24,全校物理考试获得二等奖的有 $1000 \times 0.24 = 240$ 人,故②正确;
该考场化学考试被淘汰的考生的频率为0.38,采用分层随机成绩为从全校抽取200人,则化学考试成绩为淘汰的有 $200 \times 0.38 = 76$ 人,故③错误.综上所述,C正确.

- 14. A** 【解析】由题图②知,小张一星期的食品开支为 $30+40+100+80+50=300$ 元,其中肉类开支为100元,占食品开支的 $\frac{1}{3}$,而食品开支占总开支的30%,则小张一星期的肉类开支占总开支的百分比为 $30\% \times \frac{1}{3} = 10\%$.故A正确.

7. 重难上分

- 1. D** 【解析】根据题意设2020年总收入为 a ,则2021年总收入为 $2a$,2020年甲系列产品收入为 $40\%a = 0.4a$,2021年甲系列产品收入为 $20\% \cdot 2a = 0.4a$,故A正确;
2020年丁系列产品收入为15%

$a = 0.15a$,2021年丁系列产品收入为 $30\% \cdot 2a = 0.6a$,故B正确;
2021年丙和丁系列产品收入之和为 $(25\%+30\%) \cdot 2a = 1.1a$,2020年总收入为 a ,故C正确;
2020年的乙和丙系列产品收入之和的2倍为 $2(20\%+10\%) \cdot a = 0.6a$,2021年的乙和丙系列产品收入之和为 $(20\%+25\%) \cdot 2a = 0.9a$,故D错误.

- 2. C** 【解析】根据图①可知,2017~2021年全国居民人均可支配收入逐年递增,故A正确,C错误;
根据图②可知,2021年全国居民人均消费支出构成中教育文化娱乐占比为10.8%,医疗保健占比为8.8%,故B正确;
根据图②可知,2021年全国居民人均消费支出构成中食品烟酒和居住占比分别为29.8%,23.4%, $29.8\%+23.4\% = 53.2\% > 50\%$,故D正确.
- 3. D** 【解析】由题意知,该市老年低收入家庭共有900户,所占比例为6%,则该市共有低收入家庭 $900 \div 6\% = 15000$ 户,故A正确;
该市从业人员中,低收入家庭共有 $15000 \times 12\% = 1800$ 户,故B正确;
该市无业人员中,低收入家庭有 $15000 \times 29\% = 4350$ 户,故C正确;
该市大于18岁在读学生中,低收入家庭有 $15000 \times 4\% = 600$ 户,故D错误.

- 4. A** 【解析】设新农村建设前农村经济收入为 a ,可得新农村建设后农村的经济收入为 $2a$,则新农村建设前,农村的种植收入为 $0.6a$,其他收入为 $0.04a$,养殖收入为 $0.3a$,第三产业收入为 $0.06a$,新农村建设后,农村的种植收入为

0.74a, 其他收入为 0.1a, 养殖收入为 0.6a, 第三产业收入为 0.56a, 新农村建设后, 种植收入增加, 其他收入增加了一倍以上, 养殖收入增加了一倍, 养殖收入与第三产业收入的总和占总收入的比例为 $30\% + 28\% = 58\% > 0.5$, 超过经济收入的一半, 故 A 错误, B, C, D 正确.

5. C 【解析】由题图可知, 甲企业月收入数据比乙企业月收入数据都高, 故 A 正确;
由题图可知, 甲、乙两企业月收入差距如下:

月份	1	2	3	4	5
差距/万元	200	300	200	100	300
月份	6	7	8	9	10
差距/万元	300	600	400	300	300

即差距的最大值在 7 月份, 为 600 万元, 故 B 正确;

由上表可知, 甲、乙两企业月收入差距的平均值为:

$$\frac{1}{10} \times (200 + 300 + 200 + 100 + 300 +$$

$300 + 600 + 400 + 300 + 300) = 300$ 万元, 故 C 错误;

\therefore 10 月份与 6 月份相比, 甲企业与乙企业的月收入都增加了 200 万元, 但甲企业 6 月份的收入为 600 万元, 乙企业 6 月份的收入为 300 万元, \therefore 甲企业月收入的增长率比乙企业月收入的增长率低, 故 D 正确.

6. ACD 【解析】由题图知, 2016—2020 年, 我国快递业务量持续增长, 故 A 正确;

2016—2019 年, 我国快递业务量增长速度持续下降, 但 2019—2020 年, 我国快递业务量增长速度上升, 故 B 错误;

2017—2020 年, 我国快递业务量持续增长, 且保持在较高水平, 可以估计我国 2021 年快递业务量持续增长, 故 C 正确;

设我国 2015 年的快递业务量为 x 亿件, 则 $(1 + 51.4\%)x = 312.8$, $x = \frac{312.8}{1.514} \approx 206.6 < 210$, 故 D 正确.

7. ACD 【解析】从题图甲可知, 2022 年城镇居民人均可支配收入增长额为 $49\,283 - 47\,412 = 1\,871$, 2022 年农村居民人均可支配收入增长额为 $20\,133 - 18\,931 = 1\,202$, 故 A 正确;

从题图甲可知, 2022 年城镇居民收入实际增速为 1.9%, 2022 年农村居民收入实际增速为 4.2%, 故 B 错误;

从题图乙可知, 2022 年食品支出总额占个人消费支出总额的比重为 30.5%, 属于 30%~40% 的范围, 故 C 正确;

从题图乙可知, 2022 年食品烟酒和居住占比为 $30.5\% + 24.0\% = 54.5\%$, 故 D 正确.

S3 考点训练

1. B 【解析】依题意, 全球年平均气温在区间 $[14.35, 14.75]$ 内的频率是 $0.2 \times 0.5 + 0.2 \times 0.65 = 0.23$, 故全球年平均气温在区间 $[14.35, 14.75]$ 内的有 $100 \times 0.23 = 23$ 年. 故 B 正确.

2. C 【解析】由题图得高于 60 分的频率为 $(0.020 + 0.015) \times 20 = 0.7$. \therefore 高于 60 分的人数是 350, \therefore 高一学生人数是 $\frac{350}{0.7} = 500$. 故 C 正确.

3. D 【解析】这 5 年中, 销售额单调递增, 故 A 错误;
增长率先变大后变小再变大, 故 B 错误;

2018 年的增长率最大, 故 C 错误;
2021 年销售额最大, 故 D 正确.

4. 0.005 2 【解析】由题图知, $(2 \times a + 0.02 + 0.025 + 0.045) \times 10 = 1$, 解得 $a = 0.005$.

因为在 $[50, 60]$ 内和 $[60, 70]$ 内的样本个数比例为 $0.020 : 0.005 = 4 : 1$, 根据分层随机抽样可知, 日支出在 $[60, 70]$ 中被抽取的人数为 $10 \times \frac{1}{1+4} = 2$.

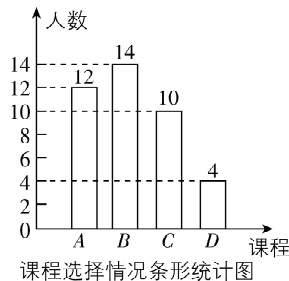
5. 140 【解析】由题意得, 这 200 名学生中每周的自习时间不少于 22.5 小时的人数为

$$200 \times (0.16 + 0.08 + 0.04) \times 2.5 = 140.$$

6. 【解】(1) $12 \div 30\% = 40$ 人, 故这次学校抽查的学生人数是 40 人.

(2) 样本中选择 C 课程的人数为 $40 - 12 - 14 - 4 = 10$.

补全条形统计图如下:



(3) $360^\circ \times \frac{14}{40} = 126^\circ$, 即 B 课程部分的圆心角的度数为 126° .

(4) $1\,000 \times \frac{4}{40} = 100$ 人, 即该校 1 000 名学生中报 D 选修课的大约有 100 人.

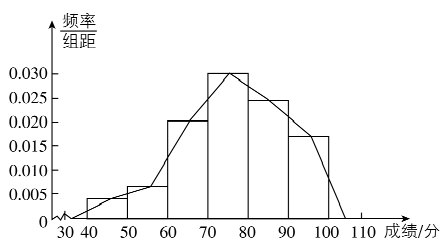
7. 【解】(1) 频率分布表如下:

分组	频数	频率
$[40, 50)$	2	0.04
$[50, 60)$	3	0.06
$[60, 70)$	10	0.20
$[70, 80)$	15	0.30
$[80, 90)$	12	0.24

续表

分组	频数	频率
[90,100]	8	0.16
合计	50	1

(2) 频率分布直方图及频率折线图如图所示:



8. 【解】(1) 由题图可得 $(0.01 + 0.015 + 0.015 + m + 0.025 + 0.005) \times 10 = 1$, 解得 $m = 0.030$.

(2) 由题中频率分布直方图可知, 100 个产品中, 一等品、二等品分别有 60 个、40 个, 故从一等品中抽取 $\frac{60}{100} \times 5 = 3$ 个, 从二等品中抽取 $\frac{40}{100} \times 5 = 2$ 个.

§4 用样本估计总体的数字特征

基础满分

1. B 【解析】根据题表得 $\frac{10 \times 5 + 15 \times 6 + 20 \times 7 + 5 \times 8}{50} = 6.4$, 则

这 50 名学生这一周在校的平均体育锻炼时间是 6.4 小时. 故 B 正确.

2. D 【解析】将数据由小到大进行排列得 5, 5, 6, 7, 8, 8, 8, 因此, 这组数据的众数为 8, 中位数为 7. 故 D 正确.

3. D 【解析】由题意可得, $\frac{6+7+8+a+10}{5} = 8$, 解得 $a = 9$. 故 D 正确.

4. C 【解析】根据题意, 100 个数据

的中位数是 8, 而 100 为偶数, 所以第 50 个和第 51 个数据的平均数为中位数, 即第 50 个和第 51 个数据的平均数为 8, 故 C 正确, A, B 不一定正确, D 错误.

5. C 【解析】根据题意, 若 $a^2 - 9b - 4 = 0$, 则有 $a^2 - 4 = 9b$, 即 $(a+2)(a-2) = 9b$. 由于 b 是质数, 故 $9b$ 的因数只可能是 1, 3, 9, b , $3b$, $9b$, 注意到 a 是质数, 故 $a \geq 2$, 则 $a+2 \geq 4$, 故 $a+2$ 可能的取值是 9, b , $3b$, $9b$, 于是可能有以下四种情况:

$$\textcircled{1} \begin{cases} a+2=9, \\ a-2=b, \end{cases} \text{解得} \begin{cases} a=7, \\ b=5; \end{cases}$$

$$\textcircled{2} \begin{cases} a+2=b, \\ a-2=9, \end{cases} \text{解得} \begin{cases} a=11, \\ b=13; \end{cases}$$

$$\textcircled{3} \begin{cases} a+2=3b, \\ a-2=3, \end{cases} \text{无解;} \quad \textcircled{4} \begin{cases} a+2=9b, \\ a-2=1, \end{cases} \text{无解.}$$

当 $\begin{cases} a=7, \\ b=5 \end{cases}$ 时, 2, 3, 5, 7 的中位数是

$$\frac{3+5}{2} = 4;$$

当 $\begin{cases} a=11, \\ b=13 \end{cases}$ 时, 2, 3, 11, 13 的中位数

是 $\frac{3+11}{2} = 7$. 故 C 正确.

6. 61 【解析】 \because 数据 x_1, x_2, \dots, x_n 的平均数为 30, \therefore 数据 $2x_1+1, 2x_2+1, \dots, 2x_n+1$ 的平均数为 $2 \times 30+1=61$.

7. C 【解析】这 4 个数中最大的数是 7, 故 A 错误;

这 4 个数的极差是 $7 - (-9) = 16$, 故 B 错误;

4 个数从小到大排列为 -9, 2, 3, 7, 中位数为 $\frac{1}{2} \times (2+3) = 2.5$,

故 C 正确; 由于不知道这一个月中其他日期的气温, 无法确定这一

个月的平均气温, 故 D 错误.

8. D 【解析】因为数据 x_1, x_2, \dots, x_n 的平均数为 2, 方差为 1, 所以 $3x_1, 3x_2, \dots, 3x_n$ 的平均数为 6, 方差为 $3^2 \times 1 = 9$, 所以 $3x_1+2, 3x_2+2, \dots, 3x_n+2$ 的平均数为 8, 方差为 9. 故 D 正确.

9. A 【解析】由题中折线统计图可知, 甲、丙的平均数的水平线高于乙、丁的平均数水平线, 即甲、丙的成绩相对较好;

比较题中甲、丙的折线图可知, 甲的成绩相对于平均成绩的波动幅度小于丙的成绩相对于平均成绩的波动幅度, 即甲的成绩更稳定. 所以这四人中甲的成绩好又发挥稳定. 故 A 正确.

10. AC 【解析】因为 $x_1+x_2+x_3+x_4=200, x_5=10, x_6=90$, 所以 $\bar{x} = \frac{200+10+90}{6} = 50$, 故 A 正确, B 错误;

因为 $x_i \in (10, 90), i=1, 2, 3, 4$, 所以去除 $x_5=10, x_6=90$ 后, 剩余数据的平均数不变, 则剩余数据的波动性更小, 方差更小, 所以 $m > n$, 故 C 正确, D 错误.

11. 20 【解析】由题意得, 标准差 $s = 5\sqrt{2}$, 则 $s^2 = 50$, 即 $s^2 = \frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x})^2 = \frac{1}{10} \left(\sum_{i=1}^{10} x_i^2 - 10\bar{x}^2 \right) = 50$, 代入数据得 $\frac{1}{10} (4\,500 - 10\bar{x}^2) = 50$, 解得 $\bar{x} = 20$.

12. $s_1^2 > s_2^2$ 【解析】 $\bar{x}_1 = \frac{1}{6} [-5+7+15+14+(-4)+(-3)] = 4, s_1^2 = \frac{1}{6} \times [(-5-4)^2 + (7-4)^2 + (15-4)^2 + (14-4)^2 + (-4-4)^2 + (-3-4)^2] \approx 70.67;$

$$\bar{x}_2 = \frac{1+4+10+7+2+0}{6} = 4, s_2^2 =$$

$$\frac{1}{6} \times [(1-4)^2 + (4-4)^2 + (10-4)^2 + (7-4)^2 + (2-4)^2 + (0-4)^2] \approx 12.34. \text{ 故 } s_1^2 > s_2^2.$$

【一题多解】由表中数据可得,甲地温度的波动明显大于乙地的波动,故 $s_1^2 > s_2^2$. 故答案为 $s_1^2 > s_2^2$.



重难上分

1. D 【解析】由题意将出生率从小到大排列为 8.52‰, 10.41‰, 10.86‰, 11.99‰, 12.64‰, 13.03‰, 13.27‰, 13.57‰, 13.83‰, 14.57‰, 而 $10 \times 60\% = 6$, 所以这组数据的 60% 分位数为第 6 个和第 7 个数据的平均数, 即 $(13.03\% + 13.27\%) \times \frac{1}{2} = 13.15\%$. 故 D 正确.

2. A 【解析】根据题意, 甲组数据的 25% 分位数为 14, $8 \times 25\% = 2$, 则有 $9 < x < 16$ 且 $\frac{1}{2}(x+16) = 14$, 解得 $x = 12$.

乙组数据的平均数为 18.5, 则有 $\frac{1}{8} \times (8 + 17 + y + 13 + 24 + 28 + 20 + 22) = 18.5$, 解得 $y = 16$. 故 A 正确.

3. ABC 【解析】将数据从小到大排列得 46, 48, 51, 53, 53, 56, 56, 56, 58, 64, 66, 71 共 12 个数据, 因为 $12 \times 20\% = 2.4$, 所以 20% 分位数为第三个数据, 即为 51, 故 A 正确; 因为 $12 \times 75\% = 9$, 所以 75% 分位数为 $\frac{58+64}{2} = 61$, 故 B 正确;

该组数据的中位数为 $\frac{56+56}{2} = 56$, 故 C 正确;

该组数据的平均数为 $\frac{1}{12} \times (46+48+$

$51+53+53+56+56+56+58+64+66+71) = 56.5$, 故 D 错误.

4. 48 【解析】5 月 1 日到 10 日 PM2.5 日均值从小到大依次为 30, 32, 34, 40, 41, 45, 48, 60, 78, 80, 由于 $10 \times 64\% = 6.4$, 故这组数据的 64% 分位数为第七个数据 48.

5. 【解】(1) 这 40 个用户评分按从小到大排列如下: 63, 66, 72, 73, 74, 74, 75, 76, 76, 76, 77, 78, 78, 78, 79, 79, 80, 81, 81, 81, 82, 82, 83, 83, 84, 84, 85, 85, 86, 86, 88, 88, 89, 89, 91, 92, 93, 95, 95, 97, 得到 $25\% \times 40 = 10$, $95\% \times 40 = 38$, 可知这 40 个用户评分的 25%, 95% 分位数分别为第 10 项和第 11 项数据的平均数, 第 38 项和第 39 项数据的平均数, 分别为 76.5, 95, 据此估计该地区所有用户评分的 25%, 95% 分位数分别约为 76.5, 95.

$$(2) \bar{x} = \frac{1}{10} \times (92 + 84 + 86 + 78 + 89 + 74 + 83 + 78 + 77 + 89) = 83;$$

$$s^2 = \frac{1}{10} \times [(92-83)^2 + (84-83)^2 + (86-83)^2 + (78-83)^2 + (89-83)^2 + (74-83)^2 + (83-83)^2 + (78-83)^2 + (77-83)^2 + (89-83)^2] = 33.$$

(3) 由题意知评分在 $(83 - \sqrt{33}, 83 + \sqrt{33})$, 即 $(77.26, 88.74)$ 内的满意度等级为“A 级”, 样本中评分在 $(77.26, 88.74)$ 内的有 5 人, 则可估计该地区满意度等级为“A 级”的用户所占的百分比约为 $\frac{5}{10} \times 100\% = 50\%$.

6. A 【解析】因为在频率分布直方图中, 中位数两侧的小矩形的面积相等, 设中位数为 x , 又 $(0.04 + 0.12) \times 2 = 0.32 < 0.5$, 所以 $0.04 \times 2 + 0.12 \times 2 + (x-6) \times 0.15 = 0.5$, 可解

出 $x = 7.2$. 故 A 正确.

7. D 【解析】由题图可得 $(0.005 + x + 0.040 + 0.030 + 0.005) \times 10 = 1$, 解得 $x = 0.020$, 故 A 正确;

由题图可知, 得分在 70 分以上的人数为 $(0.040 + 0.030 + 0.005) \times 10 \times 100 = 75$, 故 B 正确;

由题图可知, 这组数据平均数的估计值为 $55 \times 0.005 \times 10 + 65 \times 0.020 \times 10 + 75 \times 0.040 \times 10 + 85 \times 0.030 \times 10 + 95 \times 0.005 \times 10 = 76$, 故 C 正确;

由题图可知, 在 $[50, 70)$ 内的频率为 $0.005 \times 10 + 0.020 \times 10 = 0.25 < 0.50$, 在 $[50, 80)$ 内的频率为 $0.005 \times 10 + 0.020 \times 10 + 0.04 \times 10 = 0.65 > 0.5$, 则中位数在 $[70, 80)$ 内, 设中位数的估计值为 x , 则 $0.005 \times 10 + 0.020 \times 10 + (x-70) \times 0.04 = 0.5$, 解得 $x = 76.25$, 故 D 错误.

8. 14 【解析】由题图可得, 第一组的频率为 $0.04 \times 5 = 0.2 < 0.6$, 前两组的频率之和为 $0.04 \times 5 + 0.1 \times 5 = 0.7 > 0.6$, 则 60% 分位数在 $[10, 15)$ 内, 设其为 x , 则 $0.1 \times (x-10) = 0.6 - 0.2$, 解得 $x = 14$. 即这组数据的 60% 分位数为 14.

9. 【解】(1) 由题图可得, $(m + 5m + 0.03 + 0.04) \times 10 = 1$, 解得 $m = 0.005$.

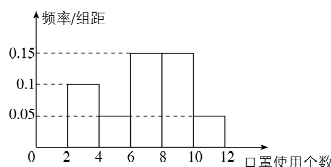
(2) 估计这 100 名学生竞赛成绩的平均数

$$\bar{x} = 65 \times 0.005 \times 10 + 75 \times 0.025 \times 10 + 85 \times 0.04 \times 10 + 95 \times 0.03 \times 10 = 84.5.$$

(3) 设被嘉奖的学生的分数不低于 x 分, 因为第四组的频率为 $0.03 \times 10 = 0.3$, 第三组的频率为 $0.04 \times 10 = 0.4$, 所以 $x \in [80, 90)$, 所以 $0.04 \times (90-x) + 0.3 = 0.4$, 解得 $x = 87.5$, 即被嘉奖的学生的分数应不低于 87.5 分.

10. 【解】(1) 由已知得 $n = \frac{14\ 000}{20\ 000} - 0.3 - 0.1 = 0.3$, $m = \frac{6\ 000}{20\ 000} - 0.2 = 0.1$.

补全频率分布直方图如图.



(2) 在 $[2, 6)$ 内的频率为 $(0.1 + 0.05) \times 2 = 0.3$, 在 $[2, 8)$ 内的频率为 $(0.1 + 0.05 + 0.15) \times 2 = 0.6$, 在 $[2, 10)$ 内的频率为 $(0.1 + 0.05 + 0.15 + 0.15) \times 2 = 0.9$, 因此估计 A 地区居民一周口罩使用个数的 75% 分位数为 $8 + \frac{0.75 - 0.6}{0.3} \times 2 = 9$, 估计 A 地区居民一周口罩使用个数的中位数为 $6 + \frac{0.5 - 0.3}{0.3} \times 2 \approx 7.3$.

(3) 由频率分布直方图得一周内使用口罩的平均数 $\bar{x} = 3 \times 0.2 + 5 \times 0.1 + 7 \times 0.3 + 9 \times 0.3 + 11 \times 0.1 = 7$;

方差 $s^2 = 0.2 \times (3-7)^2 + 0.1 \times (5-7)^2 + 0.3 \times (7-7)^2 + 0.3 \times (9-7)^2 + 0.1 \times (11-7)^2 = 6.4$.

11. D 【解析】高一年级有女生 504 人, 男生 596 人, 总人数为 $504 + 596 = 1\ 100$, 从高一女生和男生中随机抽取 50 人和 60 人, 没有按照比例分配的方式进行抽样, 不能直接用样本平均数估计总体平均数, 需要按照抽取的女生和男生在总人数中的比例计算总体的平均体重, 即 $\frac{504}{1\ 100} \times 49 + \frac{596}{1\ 100} \times 57$. 故 D 正确.

12. A 【解析】由已知可得, 骑自行车平均用时(分钟): $\bar{x}_1 = 30$, 方差 $s_1^2 = 36$;

骑电动车平均用时(分钟): $\bar{x}_2 = 20$, 方差 $s_2^2 = 16$.

骑自行车人数占总人数的 $\frac{2}{5}$, 骑

电动车人数占总人数的 $\frac{3}{5}$, 则这

50 名学生到校时间的平均数 $\bar{x} =$

$\frac{2}{5} \times 30 + \frac{3}{5} \times 20 = 24$, 方差

$s^2 = \frac{2}{5} \times [36 + (30-24)^2] + \frac{3}{5} \times$

$[16 + (20-24)^2] = 48$. 故 A 正确.

13. C 【解析】甲班与乙班的学生人数之比为 2 : 3, 则甲班 10 名学生, 乙班 15 名学生.

由题意可知这 25 名学生成绩的平均数为 $\frac{90 \times 10 + 85 \times 15}{25} = 87$,

这 25 名学生成绩的方差为 $s^2 = \frac{10}{25} \times [3 + (90-87)^2] + \frac{15}{25} \times [5 + (85-87)^2] = 10.2$. 故 C 正确.

14. $\frac{11}{9}$ 【解析】由题意可知, 甲、乙

两同学抽取的样本合在一起组成一个容量为 30 的样本, 该样本平

均数 $\bar{x} = \frac{20 \times 5 + 10 \times 4}{20 + 10} = \frac{14}{3}$, 故合在

一起后的样本方差 $s^2 = \frac{20}{30} \times$

$\left[1 + \left(5 - \frac{14}{3}\right)^2\right] + \frac{10}{30} \times \left[1 + \left(4 - \frac{14}{3}\right)^2\right] = \frac{11}{9}$.

S4 考点训练

1. D 【解析】 \because 样本 $a, 0, 1, 2, 3$ 的平均值为 1, $\therefore \frac{a+6}{5} = 1$, 解得 $a = -1$, 则样本的方差 $s^2 = \frac{1}{5} \times [(-1-1)^2 + (0-1)^2 + (1-1)^2 + (2-1)^2 + (3-1)^2] = 2$, 所以标准差 $s = \sqrt{s^2} = \sqrt{2}$. 故 D 正确.

2. C 【解析】由题意可得, 数据从小

到大排列为 23, 24, 24, 25, 26, 26, 26, 27, 29, 31, 而 $10 \times 40\% = 4$, 则这组数据的 40% 分位数为 $\frac{25+26}{2} =$

25.5. 故 C 正确.

3. C 【解析】极差为数据的最大值与最小值之差, 则极差为 $84 - 69 = 15$, 故 A 错误;

平均数为 $\frac{1}{7} \times (83 + 84 + 80 + 69 + 82 + 81 + 81) = 80$, 故 C 正确;

方差为 $s^2 = \frac{1}{7} [(83-80)^2 + (84-80)^2 + (80-80)^2 + (69-80)^2 + (82-80)^2 + (81-80)^2 + (81-80)^2] = \frac{152}{7}$, 故 B 错误;

数据由小到大排列为 69, 80, 81, 81, 82, 83, 84, 则中位数为 81, 故 D 错误.

4. D 【解析】设样本数据 x_1, x_2, \dots, x_n 的平均数为 \bar{x} , 中位数为 A , 则新样本数据 y_1, y_2, \dots, y_n 的平均数是 $\bar{x} + c$, 中位数是 $A + c$, 平均数与中位数都不同, 故 A, B 错误;

由样本数据 x_1, x_2, \dots, x_n , 得到新样本数据 y_1, y_2, \dots, y_n , 两组数据的波动性相同, 所以标准差相同, 故 C 错误;

设样本数据 x_1, x_2, \dots, x_n 的最大值为 x_1 , 最小值为 x_2 , 则新样本数据 y_1, y_2, \dots, y_n 的最大值为 y_1 , 最小值为 y_2 , 其极差不变, 都是 $x_1 - x_2$, 故 D 正确.

5. 5 2 【解析】根据题意, 设其他的 8 个数据依次为 $a_1, a_2, a_3, \dots, a_8$.

原来容量为 9 的样本的平均数为 $\frac{44}{9}$, 方差为 $\frac{152}{81}$, 则有 $\bar{x} = \frac{1}{9} (a_1 + a_2 +$

$a_3 + \dots + a_8 + 4) = \frac{44}{9}$, 则 $(a_1 + a_2 +$

$a_3 + \dots + a_8) = 40$;

$$s^2 = \frac{1}{9} [a_1^2 + a_2^2 + a_3^2 + \cdots + a_8^2 + 4^2 - 9 \times \left(\frac{44}{9}\right)^2] = \frac{152}{81}, \text{ 则 } (a_1^2 + a_2^2 + a_3^2 + \cdots + a_8^2) = 216.$$

把这个样本中一个为 4 的数据去掉, 变成一个容量为 8 的新样本,

$$\text{则新样本的平均数为 } \bar{x}' = \frac{1}{8} (a_1 + a_2 + a_3 + \cdots + a_8) = \frac{40}{8} = 5;$$

$$\text{方差 } s'^2 = \frac{1}{8} (a_1^2 + a_2^2 + a_3^2 + \cdots + a_8^2 - 8 \times 5^2) = 2.$$

6.6.8 【解析】由题意可得, $\bar{x}_{\text{甲}} = 12, s_{\text{甲}}^2 = 4.5, \bar{x}_{\text{乙}} = 16, s_{\text{乙}}^2 = 3.5$, 则这种零件尺寸的平均数估计值为 $\bar{x} = \frac{10}{10+30} \times 12 + \frac{30}{10+30} \times 16 = 15$, 方差的估计值为 $s^2 = \frac{10}{10+30} \times [4.5 + (12-15)^2] + \frac{30}{10+30} \times [3.5 + (16-15)^2] = 6.75 \approx 6.8$.

7.-1 【解析】设新数据 $3x_1+1, 3x_2+1, 3x_3+1, \cdots, 3x_n+1$ 的平均数为 \bar{x}_1 , 方差为 s_1^2 , \because 数据 $x_1, x_2, x_3, \cdots, x_n$ 的平均数为 \bar{x} , 方差为 s^2 , $\therefore \bar{x}_1 = 3\bar{x}+1, s_1^2 = 9s^2$, $\therefore 3x_1+1, 3x_2+1, 3x_3+1, \cdots, 3x_n+1$ 的平均数比方差大 4, $\therefore 3\bar{x}+1 = 9s^2+4$, $\therefore s^2 = \frac{1}{3}\bar{x} - \frac{1}{3}$, $\therefore s^2 - \bar{x}^2 = \frac{1}{3}\bar{x} - \frac{1}{3} - \bar{x}^2 = -\left(\bar{x} - \frac{1}{6}\right)^2 - \frac{11}{36}$, 又 $\because s^2 = \frac{1}{3}\bar{x} - \frac{1}{3} \geq 0$, $\therefore \bar{x} \geq 1$, 故当 $\bar{x} = 1$ 时, $s^2 - \bar{x}^2$ 取得最大值, 最大值为 -1.

8. 【解】(1) 由题意知, $m+n = 50-6-8-20 = 16, 69.2m+69.4n = 69 \times 50-$

$$68.6 \times 6 - 68.8 \times 8 - 69 \times 20 = 1108,$$

解得 $m=12, n=4$.

(2) 甲工厂样本的平均值为 69, 乙工厂样本的平均值为 $\bar{x} = \frac{1}{50} \times (68.6 \times 5 + 68.8 \times 5 + 69 \times 26 + 69.2 \times 13 + 69.4 \times 1) = 69$;

$$\text{甲工厂样本的方差为 } s_{\text{甲}}^2 = \frac{1}{50} [(68.6-69)^2 \times 6 + (68.8-69)^2 \times 8 + (69-69)^2 \times 20 + (69.2-69)^2 \times 12 + (69.4-69)^2 \times 4] = 0.048;$$

$$\text{乙工厂样本的方差为 } s_{\text{乙}}^2 = \frac{1}{50} [(68.6-69)^2 \times 5 + (68.8-69)^2 \times 5 + (69-69)^2 \times 26 + (69.2-69)^2 \times 13 + (69.4-69)^2 \times 1] = 0.0336.$$

甲、乙两个工厂生产的足球的周长平均值相同, 乙的方差较小些, 生产的足球的大小更加稳定, 所以乙工厂的生产水平更高一些.

9. 【解】(1) 质量指标值的样本平均数为 $80 \times 0.06 + 90 \times 0.26 + 100 \times 0.38 + 110 \times 0.22 + 120 \times 0.08 = 100$. 质量指标值的样本众数为 $\frac{95+105}{2} = 100$.

所以这种产品质量指标值的平均数约为 100, 众数约为 100.

综上可知中位数落在 $[95, 105)$ 内, 设中位数为 x , 则 $0.006 \times 10 + 0.026 \times 10 + 0.038 \times (x-95) = 0.5$, 解得 $x \approx 99.7$, 因此, 中位数约为 99.7.

(2) 质量指标值不低于 95 的产品所占比例约为 $0.38+0.22+0.08 = 0.68$, 由于该估计值小于 0.8, 故不能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 80%”的规定. 因为 $0.38+0.22+0.08 = 0.68$ 离 0.8 还差 0.12, $0.12 = 0.026y, y \approx 4.6$, 故质量指标值应定为不低于 $95-4.6 = 90.4$.

10. 【解】(1) 由题中的数据可得,

$$\bar{x} = \frac{1}{10} \times (9.8 + 10.3 + 10.0 + 10.2 + 9.9 + 9.8 + 10.0 + 10.1 + 10.2 + 9.7) = 10;$$

$$\bar{y} = \frac{1}{10} \times (10.1 + 10.4 + 10.1 + 10.0 + 10.1 + 10.3 + 10.6 + 10.5 + 10.4 + 10.5) = 10.3;$$

$$s_1^2 = \frac{1}{10} \times [(9.8-10)^2 + (10.3-10)^2 + (10-10)^2 + (10.2-10)^2 + (9.9-10)^2 + (9.8-10)^2 + (10-10)^2 + (10.1-10)^2 + (10.2-10)^2 + (9.7-10)^2] = 0.036;$$

$$s_2^2 = \frac{1}{10} \times [(10.1-10.3)^2 + (10.4-10.3)^2 + (10.1-10.3)^2 + (10.0-10.3)^2 + (10.1-10.3)^2 + (10.3-10.3)^2 + (10.6-10.3)^2 + (10.5-10.3)^2 + (10.4-10.3)^2 + (10.5-10.3)^2] = 0.04.$$

$$(2) \bar{y} - \bar{x} = 10.3 - 10 = 0.3, \frac{s_1^2 + s_2^2}{10} =$$

$$\frac{0.036 + 0.04}{10} = 0.0076. \text{ 因为}$$

$$\left(\frac{\bar{y} - \bar{x}}{2}\right)^2 = 0.0225 > 0.0076,$$

所以 $\bar{y} - \bar{x} > 2\sqrt{\frac{s_1^2 + s_2^2}{10}}$, 故可以认为新设备生产产品的该项指标的均值较旧设备有显著提高.

第六章 综合检测

1. C 【解析】调查某池塘中现有鱼的数量,应采用抽样调查,故 A 错误;

调查某批次汽车的抗撞击能力,应采用抽样调查,故 B 错误;

选出某班短跑最快的学生参加全校短跑比赛,适宜采用普查,故 C 正确;

调查市场上某种食品的色素含量是否符合国家标准,应采用抽样调查,故 D 错误.

2. D 【解析】总体是全校 3 000 名学生的身高,故 A 错误;

个体是全校 3 000 名学生中每名学生的身体,故 B 错误;

样本是抽取到的 100 名学生的身高,故 C 错误;

样本容量是 100,故 D 正确.

3. D 【解析】分层随机抽样是从一个可以分成不同子总体(或称为层)的总体中,按规定的比例从不同层中随机抽取样本(个体)的方法. 分层随机抽样必须保证每个个体被抽到的可能性相等,这个可能性等于样本容量与总体个数的比. 由以上的论述,可知每层的个体数可能不相等,每层中抽取的个体数也可能不相等,但不同的层中,每个个体被抽到的可能性必须相等. 故 ABC 错误,D 正确.

4. B 【解析】由题意可知,读取的标号依次为 175,331,572(舍去),455,068, ..., 故第三袋牛奶的标号是 455. 故 B 正确.

5. B 【解析】因为所研究的总体中个体差异很大,所以为了让样本具有代表性,最合理的抽样方法是分

层随机抽样. 故 B 正确.

6. C 【解析】由题意可得,这组数据从小到大排列为 4,5,5,6,7,8,9, $7 \times 75\% = 5.25$, 所以这组数据的 75% 分位数为 8. 故 C 正确.

7. D 【解析】因为 $80+70=60+90$, 则在更正数据之后,数据的平均数不变,即平均数 $\bar{x}=70$; 不妨设其他 48 个数据依次为 a_1, a_2, \dots, a_{48} , 则 $(a_1-70)^2 + (a_2-70)^2 + \dots + (a_{48}-70)^2 + (60-70)^2 + (90-70)^2 = 50 \times 75, (a_1-70)^2 + (a_2-70)^2 + \dots + (a_{48}-70)^2 + (80-70)^2 + (70-70)^2 = 50s^2$, 即 $50(s^2-75) = 100-400-100 = -400 < 0$, 则 $s^2 < 75$. 故 D 正确.

8. D 【解析】1 班学生得分的平均分为

$$\frac{4 \times 4 + 20 \times 6 + 6 \times 7 + 4 \times 8 + 4 \times 9 + 2 \times 10}{4 + 20 + 6 + 4 + 4 + 2} =$$

$$6.65,$$

2 班学生得分的平均分为

$$\frac{6 \times 5 + 4 \times 6 + 8 \times 7 + 16 \times 8 + 4 \times 9 + 2 \times 10}{6 + 4 + 8 + 16 + 4 + 2} =$$

$$7.35, 7.35 > 6.65, \text{故①错误;}$$

$$1 \text{ 班学生得分的方差为 } \frac{1}{40} \times [4 \times$$

$$(4-6.65)^2 + 20 \times (6-6.65)^2 + 6 \times (7-6.65)^2 + 4 \times (8-6.65)^2 + 4 \times (9-6.65)^2 + 2 \times (10-6.65)^2] =$$

$$2.2275,$$

$$2 \text{ 班学生得分的方差为 } \frac{1}{40} \times [6 \times (5-7.35)^2 + 4 \times (6-7.35)^2 + 8 \times (7-7.35)^2 + 16 \times (8-7.35)^2 + 4 \times (9-7.35)^2 + 2 \times (10-7.35)^2] =$$

$$1.8275, 1.8275 < 2.2275, \text{故}$$

②错误;
因为 $40 \times 90\% = 36$, 所以 1 班学生得分的 90% 分位数为 9, 2 班学生

得分的 90% 分位数为 9, 故③正确;

1 班学生得分的中位数为 $6 < 7$, 2 班学生得分的中位数为 $8 > 7$, 故

④正确.

综上, 故 D 正确.

9. BCD 【解析】这组数据从小到大排列为 4, 4, 5, 7, 7, 7, 8, 9, 9, 10,

计算这组数据的平均数为 $\frac{1}{10} \times (4 +$

$4 + 5 + 7 + 7 + 7 + 8 + 9 + 9 + 10) = 7$, 故 A 错误;

这组数据的众数是 7, 故 B 正确;

这组数据的极差是 $10 - 4 = 6$, 故 C 正确;

因为 $10 \times 75\% = 7.5$, 且第 8 个数是 9, 所以这组数据的 75% 分位数为 9, 故 D 正确.

10. AC 【解析】已知 $10 \times (0.01 + 0.03) = 0.4 < 0.5$, $10 \times (0.01 + 0.03 + 0.04) = 0.8 > 0.5$, 所以 50% 分位数位于区间 $[80, 90)$ 内, 设其为 x , 此时 $10 \times 0.01 + 10 \times 0.03 + (x - 80) \times 0.04 = 0.5$, 解得 $x = 82.5$, 所以该校学生测试成绩的 50% 分位数的估计值为 82.5, 故 A 正确;

由题图易知该校学生测试成绩的众数所在区间为 $[80, 90)$, 此时众数的估计值为 85, 故 B 错误;
成绩在 $[60, 70)$, $[70, 80)$, $[80, 90)$, $[90, 100]$ 的人数依次为 10, 30, 40, 20, 因为同一组中的数据用该组区间的中点值为代表, 所以 $\bar{x} =$

$$\frac{65 \times 10 + 75 \times 30 + 85 \times 40 + 95 \times 20}{100} =$$

$$82, \text{故 C 正确;}$$

已知样本中测试成绩位于 $[60, 70)$ 内的共有 10 人, 无法确定样

本中成绩位于 $[60, 64)$ 内的人数, 故无法估计该校学生测试成绩位于 $[60, 64]$ 的人数, 故 D 错误.

11. AD 【解析】因为 $a_1 < a_2 < a_3 < \dots < a_{2023}$, 该数据最中间的项为 a_{1012} , 由中位数的定义可知, $k = a_{1012}$, 故 A 正确;

不妨令 $a_n = n (n = 1, 2, \dots, 2022)$, $a_{2023} = 100^8$,

$$\text{则 } m = \frac{1+2+\dots+2022+100^8}{2023} >$$

$$\frac{1+2+\dots+2023}{2023} = 1012 = a_{1012}, \text{ 故}$$

B 错误;

不妨令 $a_n = n (n = 1, 2, \dots, 2022)$,

$$\text{且 } a_{2023} = 2022.1, \text{ 则 } m = \frac{1+2+\dots+2022+2022.1}{2023} <$$

$$\frac{1+2+\dots+2022+2023}{2023} = 1012 =$$

$a_{1012} = k$, 故 C 错误;

数据 $2a_1 + 1, 2a_2 + 1, 2a_3 + 1, \dots$,

$$2a_{2023} + 1 \text{ 的均值 } \bar{x} = \frac{\sum_{i=1}^{2023} (2a_i + 1)}{2023} =$$

$$\frac{2 \sum_{i=1}^{2023} a_i}{2023} + 1 = 2m + 1,$$

其标准差 $s_2 =$

$$\sqrt{\frac{\sum_{i=1}^{2023} [(2a_i + 1) - (2m + 1)]^2}{2023}} =$$

$$\sqrt{\frac{4 \sum_{i=1}^{2023} (a_i - m)^2}{2023}} = 2s_1, \text{ 故 D 正确.}$$

12. 85 【解析】由题可得 $\frac{70+85+t+82+75}{5} = 81$, 解得 $t = 93$, 则 5 位学生的成绩按从小到大排列为 70, 75, 82, 85, 93, $\therefore 5 \times 75\% = 3.75$, \therefore 他们成绩的 75% 分位数为 85.

13. 19 【解析】设男同学人数为 $3a$, 女同学人数为 $2a$, 则全班同

学答对题目数的平均数为 $\frac{17 \times 3a + 12 \times 2a}{3a + 2a} = 15$, 全班同学答

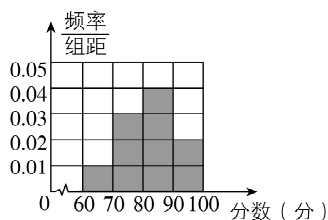
对题目数的方差为 $\frac{3}{5} \times [11 + (17-15)^2] + \frac{2}{5} \times [16 + (12-15)^2] = 19$.

14. 58.2 【解析】一组样本数据 x_1, x_2, \dots, x_{10} , 且 $x_1^2 + x_2^2 + \dots + x_{10}^2 = 2022$, 平均数 $\bar{x} = 12$, 则该组数据的方差为 $\frac{1}{10} \times [(x_1 - 12)^2 + (x_2 - 12)^2 + \dots + (x_{10} - 12)^2] = \frac{1}{10} [(x_1^2 + x_2^2 + \dots + x_{10}^2) + 10 \times 144 - 24 \times (x_1 + x_2 + \dots + x_{10})] = \frac{1}{10} \times (2022 + 1440 - 24 \times 120) = 58.2$.

15. 【解】(1) 由题意可知, 频率分布表如下:

分数	频数	频率	频率/组距
[60, 70)	10	0.1	0.01
[70, 80)	30	0.3	0.03
[80, 90)	40	0.4	0.04
[90, 100]	20	0.2	0.02

所以作出频率分布直方图如图.



(2) 由题知, 估计这 100 名员工各项素质综合评分分数的平均数 $\bar{x} = 65 \times 0.1 + 75 \times 0.3 + 85 \times 0.4 + 95 \times 0.2 = 82$, 估计这 100 名员工各项素质综合评分分数的方差 $s^2 = 0.1 \times (65 - 82)^2 + 0.3 \times (75 - 82)^2 + 0.4 \times (85 - 82)^2 + 0.2 \times (95 - 82)^2 = 81$.

(3) 因为从 100 名员工中挑选 30 名员工组建“数字乡村发展行动

部”, 又准备挑选成绩较好的员工组建“数字乡村发展行动部”, 所以应选成绩为 70% 分位数及其后的分数的员工, 设 70 百分位数为 m , 则 $0.1 + 0.3 + (m - 80) \times 0.04 = 0.7$, 解得 $m = 87.5$, 所以被挑选的员工的分数不低于 87.5 分.

16. 【解】(1) 由频率分布直方图得 $(a + 0.02 + 0.035 + 0.025 + a) \times 10 = 1$, 解得 $a = 0.01$.

(2) 估计该校学生数学成绩的平均数为

$$0.1 \times 55 + 0.2 \times 65 + 0.35 \times 75 + 0.25 \times 85 + 0.1 \times 95 = 75.5.$$

(3) 由题图易知, 估计该校学生数学成绩的 75% 分位数为 $80 + \frac{0.75 - (0.1 + 0.2 + 0.35)}{0.25} \times 10 = 84$.

17. 【解】(1) 由已知得八年级测试成绩 D 组: $85 \leq x < 90$ 的频数为 7, 又由扇形统计图知 D 组占 35%, 所以八年级进行冬奥会知识测试学生数 $n = 7 \div 35\% = 20$, 所以 $a = \frac{1}{2} \times (20 - 1 - 2 - 3 - 6) = 4$.

(2) 八年级 A, B, C 三组的频率之和为 $5\% + 5\% + 20\% = 30\% < 50\%$, 八年级 A, B, C, D 四组的频率之和为 $30\% + 35\% = 65\% > 50\%$, 所以中位数在 D 组, 将 D 组数据按从小到大排序为 85, 85, 86, 86, 87, 88, 89. 因为 $20 \times 30\% = 6$, 第 10 个与第 11 个数据分别为 86, 87, 所以中位数为 $\frac{86+87}{2} = 86.5$.

(3) 八年级 E: $90 \leq x < 95$, F: $95 \leq x < 100$ 两组占 $1 - 65\% = 35\%$, 共有 $20 \times 35\% = 7$ 人;

七年级 E: $90 \leq x < 95$, F: $95 \leq x < 100$ 两组人数共有 $3 + 1 = 4$ 人, 两个年级不低于 90 分的共有 4 +

$7=11$ 人,占样本的 $\frac{11}{40}$,所以估计该校七、八两个年级对冬奥会关注程度高的学生一共有 $\frac{11}{40} \times (500+500) = 275$ 人.

18. 【解】(1) 由题图可知, $10 \times (x + 0.015 + 0.02 + 0.03 + 0.025) = 1$, 解得 $x = 0.01$. \therefore 在 $[50, 80)$ 内的频率为 $0.1 + 0.15 + 0.2 = 0.45 < 0.6$, 在 $[50, 90)$ 内的频率为 $0.1 + 0.15 + 0.2 + 0.3 = 0.75 > 0.6$, \therefore 60% 分位数位于区间 $[80, 90)$ 内, 设为 m , 则 $0.45 + (m - 80) \times 0.03 = 0.6$, 解得 $m = 85$, \therefore 估计这 600 名学生评分的 60% 分位数为 85.

(2) \therefore 低于 80 分的学生中三组学生的人数之比为 $0.1 : 0.15 : 0.2 = 2 : 3 : 4$, \therefore 应选取评分在 $[60, 70)$ 的学生人数为 $30 \times \frac{3}{2+3+4} = 10$.

(3) 由题图知, 认可程度的平均分为 $\bar{x} = 55 \times 0.1 + 65 \times 0.15 + 75 \times 0.2 + 85 \times 0.3 + 95 \times 0.25 = 79.5 < 0.85 \times 100 = 85$, 则“校本课程”工作需要进一步整改.

19. 【解】(1) 由频率分布直方图可得, 在区间 $[0, 160]$ 的频率总和恰为 0.7, 由样本估计总体, 可得临界值 a 的值为 160, 估计众数为 $(120, 160]$ 的中间值 140, 估计平均数为 $20 \times 0.04 + 60 \times 0.12 + 100 \times 0.24 + 140 \times 0.3 + 180 \times 0.25 + 220 \times 0.05 = 130$.

(2) 由(1)知, 月用电量在 $[0, 160]$ 内的居民在使用阶梯电价前后用电量不变, 节电量为 0 千瓦·时;

月用电量在 $(160, 200]$ 内的 50 户居民, 平均每户月用电量为 180 千瓦·时, 超出部分为 20 千瓦·时, 根据题意, 每户每月节电 $20 \times 40\% = 8$ 千瓦·时, 50 户每月共节电 $8 \times 50 = 400$ 千瓦·时;

月用电量在 $(200, 240]$ 内的 10 户居民, 平均每户月用电量为 220 千瓦·时, 超出部分为 60 千瓦·时, 根据题意, 每户每月节电 $60 \times 40\% = 24$ 千瓦·时, 10 户每月共节电 $24 \times 10 = 240$ 千瓦·时;

故样本中 200 户居民每月共节电 $400 + 240 = 640$ 千瓦·时, 用样本估计总体, 得全市居民每月节电量约为 $640 \times \frac{300\,000}{200} = 96$ 万千瓦·时.

(3) 由题意可得, 全市缴纳电费总额不变, 由于“未超出部分”的用电量在“阶梯电价”前后不变, 故“超出部分”对应的总电费也不变, 在 200 户居民组成的样本中, 每月用电量共超出 $20 \times 50 + 60 \times 10 = 1\,600$ 千瓦·时, 实行“阶梯电价”后, 共节约 640 千瓦·时, 剩余 960 千瓦·时, 所以 $1\,600 \times 0.5 = 960 \times b$, 解得 $b \approx 0.83$.

第七章 概率

§1 随机现象与随机事件

基础满分

1. D 【解析】先后抛掷两枚质地均匀的硬币, 有先后顺序, 则此试验的样本空间为 $\{(\text{正面}, \text{正面}), (\text{正面}, \text{反面}), (\text{反面}, \text{正面}), (\text{反面}, \text{反面})\}$. 故 D 正确.

2. ①② 【解析】明天的事是未来才发生的事, 具有不确定性, 故①②属于随机现象; 易知在三角形中, 大角对大边, 故③属于确定性现象.

3. 10 【解析】从标有 1, 2, 3, 4, 5 的 5 张纸片中任取 2 张, 不同的取法

有 $(1, 2), (1, 3), (1, 4), (1, 5), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)$, 共 10 种.

4. 【解】(1) 依题意得, 样本空间为 $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$;

(2) 依题意得, 样本空间为 $\{(a, b), (a, c), (a, d), (a, e), (b, c), (b, d), (b, e), (c, d), (c, e), (d, e)\}$.

5. B 【解析】根据事件的分类可知, 抛硬币正面朝上, 打开电视正在播广告, 两个实数之和为正数均为随机事件, 三角形中不可能有两个直角, B 为不可能事件. 故 B 正确.

6. B 【解析】4 名男生, 2 名女生中随机抽取 3 人, 所有可能为 3 名男

生, 2 男 1 女, 1 男 2 女, 则必然事件为至少有 1 名男生. 故 B 正确.

7. BC 【解析】他投篮一次就命中为随机事件, 故 A 错误;

随机事件发生的可能性越大, 它发生的概率越接近 1, 故 B 正确;

当且仅当两枚骰子出现的点数都为 1 时, 骰子的点数和为 2, 这是有可能的, 故 C 正确;

试验“连续投掷一枚均匀的骰子直到出现 3 点时停止, 观察投掷的次数”的样本空间为 $\{1, 2, 3, 4, 5, 6, \dots\}$, 故 D 错误.

8. ACD 【解析】A, B 是两个随机事件, $A+B$ 表示两个事件至少有一个发生, 故 A 正确; $\overline{A+B}$ 表示两个事件均不发生, 故 C