



第 14 章 统计

14.1 获取数据的基本途径 及相关概念

- 1. A** 【解析】抽样调查在抽取调查对象时必须保证所抽取出的样本具有代表性,因此抽样时一定要注意事先设计好抽样的程序,按既定的程序进行抽样.
- 2. C** 【解析】A 不能用普查的方式调查,因为这种试验具有破坏性;
B 用普查的方式无法完成;
C 可以用普查的方式进行调查;
D 中试验具有破坏性,且需要耗费大量的时间,在实际生产中无法应用普查的方式调查.
故选 C.
- 3. B** 【解析】医生不可能将一个人的血液全抽出来进行检查,通常是抽取少量的血样进行检查.
- 4. C** 【解析】A,B,D 都是通过查询获取的数据,C 是通过调查获取的数据. 故选 C.
- 5. C** 【解析】调查某批次汽车的抗撞击能力,有破坏性,故用抽样调查的方式,故 A 错误;
了解全国中学生每周体育锻炼的时间,工作量大,用抽样调查的方式,故 B 错误;
为了调查市场上某种食品的色素含量是否符合国家标准,工作量大,用抽样调查的方式,故 C 正确;
对载人飞船“神舟十八号”零部件的检查十分重要,故进行普查,故 D 错误.
故选 C.
- 6. B** 【解析】对于 A,某城市元旦前后的气温,一般通过观察获取数据,故 A 错误;



对于 B, 某种新型电路元件使用寿命的测定, 一般通过试验获取数据, 故 B 正确;

对于 C, 电视台想知道某一个节目的收视率, 一般通过调查获取数据, 故 C 错误;

对于 D, 高中生日平均上网时间, 一般通过调查获取数据, 故 D 错误. 故选 B.

- 7. ABD** 【解析】了解玉米种子的发芽情况, 是具有破坏性的调查, 应采用抽样调查; 了解全班同学是否给爸妈洗过脚, 调查的对象比较少, 容易调查, 应采用普查; 了解刚生产的一批灯泡的使用寿命, 是具有破坏性的调查, 应采用抽样调查; 了解全国中学生安全自救知识掌握情况, 人数太多, 不适合普查, 应采用抽样调查.

14.2 抽样

14.2.1 简单随机抽样

- 1. A** 【解析】根据简单随机抽样的定义和特点知, ①②③④均正确. 故选 A.
- 2. CD** 【解析】对于 A: 50 名最优秀的官兵, 不具备等可能性, A 错误; 对于 B: 从中任意拿出一支铅笔检测后再放回箱子, 是有放回抽样, 不属于简单随机抽样, B 错误; 对于 C: 一次性抽取和逐个不放回地抽取等价, C 正确; 对于 D: 从 2 000 个灯泡中不放回地逐个抽取 20 个进行质量检查, 是简单随机抽样, D 正确. 故选 CD.
- 3. D** 【解析】由题意可得 $\frac{9}{n-1} = \frac{1}{4}$, 故 $n = 37$, 所以每个个体被抽到的机会为 $\frac{10}{37}$. 故选 D.
- 4. C** 【解析】由简单随机抽样的定义, 知每个个体被抽到的可能性相等, 故高一(5)班和高一(6)班被抽到的可



能性均为 $\frac{3}{10}$. 故选 C.

5. $\frac{3}{10} \quad \frac{1}{8}$ 【解析】因为简单随机抽样

过程中每个个体被抽到的可能性均为

$\frac{n}{N}$ (N 表示总体中含有的个体数, n 表

示从总体中逐个不放回抽取的个体数), 所以某一特定小球被抽到的可能

性是 $\frac{3}{10}$. 因为本题中的抽样是不放回

抽样, 所以第一次抽取时, 每个小球被

抽到的可能性为 $\frac{1}{10}$, 第二次抽取时, 剩

余 9 个小球, 每个小球被抽到的可能性

为 $\frac{1}{9}$, 第三次抽取时, 剩余 8 个小球, 每

个小球被抽到的可能性为 $\frac{1}{8}$.

6. B 【解析】由利用随机数表抽样的步骤可知各步骤先后顺序为①③②.

7. B 【解析】由随机数表法可知, 样本的前 5 个个体的编号分别为 23, 20, 26, 24, 25, 因此, 选出的第 5 个个体的编号为 25. 故选 B.

8. B 【解析】A 中总体容量较大, 样本容量也较大, 不适宜用抽签法; B 中总体容量较小, 样本容量也较小, 可用抽签法; C 中甲、乙两厂生产的两箱产品有明显区别, 不能用抽签法; D 中虽然样本容量较小, 但总体容量较大, 不适宜用抽签法. 故选 B.

9. 是 【解析】由抽签法的定义知, 题目中的抽样方法是简单随机抽样中的抽签法.

10. $0, 1, \dots, 99$ (或 $1, 2, \dots, 100$) $00, 01, \dots, 99$ (或 $001, 002, \dots, 100$)

【解析】根据两种方法的不同特点, 编号是不同的, 抽签法由小到大即可, 而随机数表法编号的位数应该相同, 所以抽签法的编号一般为 $0, 1, \dots, 99$ (或 $1, 2, \dots, 100$), 而随机数表



法的编号应为 $00, 01, \dots, 99$ (或 $001, 002, \dots, 100$).

11. 16 【解析】利用随机数表法,从第 1 行的第 5 列数字开始,从左到右依次选取两个数字,

则选出来的样本个体编号为:07, 92 (舍去), 45 (舍去), 44 (舍去), 17, 16, \dots , 所以选出来的第 3 个个体的编号为 16.

12. 44 【解析】根据随机数表的读取方法,第 2 行第 4 列的数为 3,每次从左向右选取两个数字,所得数字依次为 32, 58, 65, 74, 13, 36, 98, 32, 44. 其中 58, 65, 74, 98 不在编号范围内,舍去,再去除重复的,剩下的编号为 32, 13, 36, 44, 所以选取的第四个编号为 44.

13. 【解】第一步,确定参演学生.

(1)将 30 名高一学生从 01 到 30 编号,然后用相同的纸条做成 30 个号签,在每个号签上分别写上编号,然后放入一个暗箱中搅匀,从中顺次不放回地抽出 10 个号签,相应编号的学生参加演出;

(2)运用相同的办法分别从 18 名高二学生中抽取 6 人,从 10 名高三学生中抽取 4 人.

第二步,确定演出顺序. 确定了演出人员后,再用相同的纸条做成 20 个号签,上面分别写上 1~20 这 20 个数字,代表演出顺序,不放回地让每名学生抽一张,各人抽到的号签上的数就是这位学生表演的顺序.

14. 2. 2 分层抽样

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

2. C 【解析】因为男女生视力情况差异不大,而三个学段学生的视力情况有较大差异,所以应按学段分层抽样. 故



选 C.

3. C 【解析】 $\frac{15}{15+3+2} = \frac{30}{n}$, 解得 $n = 40$.

4. C 【解析】由题意可得 $\frac{24}{120} = \frac{k}{5+k+3}$,

解得 $k = 2$.

则 C 种型号的产品抽取的件数为

$$120 \times \frac{3}{5+2+3} = 36, \text{ 故选 C.}$$

5. B 【解析】根据分层抽样的方法, 该样本中 A 车间和 B 车间的人数共为 $300 \times 15\% + 300 \times 40\% = 165$. 故选 B.

6. BCD 【解析】因为每一种型号的轿车数量都比较多, 不适合用抽签法, 故 A 错误;

在分层抽样中, 抽样比为

$$\frac{46}{1\ 200+6\ 000+2\ 000} = \frac{1}{200}, \text{ 故 B 正确;}$$

在分层抽样中, 甲、乙、丙三种型号的轿车依次应抽取 6 辆、30 辆、10 辆, 故 C 正确;

在分层抽样中, 每一辆车被抽到的概率都是相等的, 故 D 正确. 故选 BCD.

7. C 【解析】由题意可得抽样比为

$$\frac{45-20-10}{300} = \frac{1}{20}, \text{ 所以学生总人数为 } 45 \div$$

$$\frac{1}{20} = 900, \text{ 即这个学校共有高中学生}$$

900 人. 故选 C.

8. C 【解析】因为由港珠澳大桥实现内地前往香港的老、中、青旅客的比例分别为 5:2:3, 现使用分层抽样的方法从这些旅客中随机抽取 n 人, 青年旅客

$$\text{抽到 60 人, 所以 } \frac{60}{n} = \frac{3}{5+2+3}, \text{ 解得 } n =$$

200. 则老年旅客抽到 100 人, 中年旅客抽到 40 人, 故 A, B, D 错误. 故选 C.

9. ACD 【解析】 \because 某单位共有老年人 180 人, 中年人 540 人, 青年人 a 人, 用分层抽样的方法进行调查, 样本中的

$$\text{中年人为 6 人, } \therefore m \times \frac{540}{180+540+a} = 6,$$



经检验,只有 B 选项不满足,故选 ACD.

10. 80 【解析】女学生人数所占的比例

为 $\frac{1\ 200}{300+1\ 500+1\ 200} = \frac{2}{5}$, 则应抽取

的女学生人数为 $200 \times \frac{2}{5} = 80$.

11. D 【解析】①某班元旦聚会,要产生两名“幸运者”,总体和样本容量都不大,所以采用简单随机抽样. ②总体由具有明显差异的几个部分组成,所以采用分层抽样. 故选 D.

12. ABC 【解析】在抽签法、随机数表法和分层抽样中,每个个体被抽中的概率均为 $\frac{n}{N}$, 所以 $p_1 = p_2 = p_3$.

13. BC 【解析】根据题意,选 A 的人数至少为 $2\ 017 \times 70\% = 1\ 411.9 \approx 1\ 412$ (人), 至多为 $2\ 017 \times 75\% = 1\ 512.75 \approx 1\ 512$ (人); 选 B 的人数至少为 $2\ 017 \times 40\% = 806.8 \approx 807$ (人), 至多为 $2\ 017 \times 45\% = 907.65 \approx 907$ (人), 故同时选 A, B 的人数至少为 $1\ 412 + 807 - 2\ 017 = 202$ (人), 至多为 $1\ 512 + 907 - 2\ 017 = 402$ (人), 因此符合题意的选项有 B, C. 故选 BC.

14.3 统计图表

14.3.1 扇形统计图、折线统计图、频数直方图+14.3.2 频率分布直方图

1. C 【解析】由折线统计图可以看出世界人口的总量随着年份的增加而增加, 故 A 正确;

由扇形统计图可知 2050 年亚洲人口比其他各洲人口的总和还要多, 故 B 正确;

由条形统计图可知 2050 年欧洲人口与南美洲及大洋洲人口之和基本持



平,故 D 正确;

三幅统计图并不能得到各个洲人口增长速度的快慢,故 C 错误.

故选 C.

2. BCD 【解析】A 中,该年第一季度 GDP 总量和增长率均居同一位的省份有 2 个,江苏和河南,分别居第一位和第四位,故 A 错误;

B 中,由折线统计图可得与前一年周期相比,该年第一季度五个省份的 GDP 总量均实现了增长,故 B 正确;

C 中,浙江前一年同期的 GDP 总量为 $\frac{4\,632.1}{1+3.3\%} \approx 4\,484.12$ (亿元),

江苏前一年同期的 GDP 总量为 $\frac{6\,653.2}{1+10.2\%} \approx 6\,037.39$ (亿元),

河南前一年同期的 GDP 总量为 $\frac{4\,067.4}{1+6.6\%} \approx 3\,815.57$ (亿元),

山东前一年同期的 GDP 总量为 $\frac{6\,469.3}{1+7\%} \approx 6\,046.07$ (亿元),

辽宁前一年同期的 GDP 总量为 $\frac{2\,642.2}{1+9.6\%} \approx 2\,410.77$ (亿元),

故前一年同期五个省份的 GDP 总量位于前三位的是山东、江苏、浙江,故 C 正确;

D 中,由 C 中计算可得前一年同期五个省份的 GDP 总量中,浙江的 GDP 总量也是第三位,故 D 正确.

故选 BCD.

3. D 【解析】由条形图得合唱社团人数为 70,由扇形图得合唱社团人数占比为 35%,因此选取的总人数为

$$\frac{70}{35\%} = 200,$$

由扇形图得演讲及舞蹈社团人数和占比为 $1 - 20\% - 10\% - 35\% = 35\%$,人数和为 $200 \times 35\% = 70$,

由条形图得演讲社团人数为 30,所以舞蹈社团人数为 40. 故选 D.



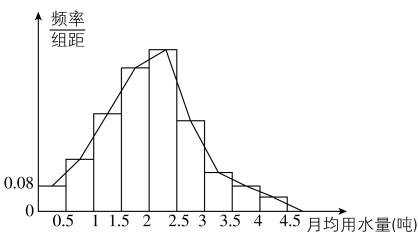
4. 600 【解析】由频率分布直方图可知, 成绩小于 60 分的学生的频率为 $10 \times (0.002 + 0.006 + 0.012) = 0.2$, 所以成绩小于 60 分的学生数是 $3\,000 \times 0.2 = 600$.

5. 【解】(1) 区间 $[0.5, 1)$ 内的频率为 0.08, 频数为 8, 区间 $[4, 4.5)$ 内的频率为 0.02, 频数为 2, 则 $x = 100 - (4 + 8 + 15 + 22 + 14 + 6 + 4 + 2) = 25$, $y = 6 \div 100 = 0.06$.

(2) 因为从左数第 4 个矩形对应的频率为 0.22, 且组距为 0.5,

所以它的高度为 $0.22 \div 0.5 = 0.44$.

(3) 由频率分布直方图, 画出频率折线图如图所示.



6. 【解】(1) 由频率分布直方图知, 样本中分数小于 50 的人数为 $100 \times [1 - (0.04 + 0.02 + 0.02 + 0.01) \times 10] = 10$.

又 \because 样本中分数小于 40 的人数为 5,

\therefore 样本中分数在 $[40, 50)$ 内的人数为

$$10 - 5 = 5, \text{ 所占频率为 } \frac{5}{100} = \frac{1}{20},$$

\therefore 估计总体中分数在区间 $[40, 50)$ 内

$$\text{的人数为 } 400 \times \frac{1}{20} = 20.$$

(2) 设样本中的男生有 x 人, 女生有 y 人.

\therefore 分数小于 70 的男生有 $\frac{x}{2}$ 人, 不小于

70 的男生有 $\frac{x}{2}$ 人, \therefore 分数不小于 70

的女生有 $\frac{x}{2}$ 人, 分数小于 70 的女生有

$$y - \frac{x}{2} \text{ 人.}$$

又分数小于 70 的总人数为 $100 \times [1 -$



$$(0.04+0.02) \times 10] = 40,$$

$$\therefore \begin{cases} x+y=100, \\ y-\frac{x}{2}+\frac{x}{2}=40, \end{cases} \quad \text{解得} \begin{cases} x=60, \\ y=40, \end{cases}$$

\therefore 样本中男女生人数之比为 $60:40=3:2$, 根据分层抽样原理可估计总体中男生和女生人数的比例为 $3:2$.

7.16 【解析】由频率分布直方图知,月收入在区间 $[1\,500, 2\,000)$ 内的频率为 $0.000\,4 \times 500 = 0.2$, 所以月收入在区间 $[1\,500, 2\,000)$ 内的居民应抽出 $80 \times 0.2 = 16$ (人).

14.4 用样本估计总体

14.4.1 用样本估计总体的集中

趋势参数+14.4.2 用样本估计

总体的离散程度参数+

14.4.3 用频率分布直方图估计

总体分布+14.4.4 百分位数

1.B 【解析】依题意这组数据一共有 5 个数,中位数为 8,则从小到大排列 8 的前面有 2 个数,后面也有 2 个数.又唯一的众数为 9,则有 2 个 9,其余数字均只出现一次,则最大数字为 9.

又极差为 3,所以最小数字为 6,所以这组数据为 6,7,8,9,9,

所以平均数为 $\frac{6+7+8+9+9}{5} = 7.8$.

2.D 【解析】这 30 名学生得分的中位

数为 $m = \frac{5+6}{2} = 5.5$, 众数为 $n = 5$,

平均数为 $\bar{x} = (3 \times 2 + 4 \times 3 + 5 \times 10 + 6 \times 6 +$

$7 \times 3 + 8 \times 2 + 9 \times 2 + 10 \times 2) \times \frac{1}{30} \approx 5.97$,

所以 $n < m < \bar{x}$. 故选 D.

3.C 【解析】由题设,甲的数据为 $\{98,$

$99, 100, 100, 100, 103\}$, 乙的数据为

$\{99, 99, 100, 100, 100, 102\}$,

所以甲的数据的平均数为 $\bar{x}_1 =$

$\frac{98+99+100+100+100+103}{6} = 100$,



乙的数据的平均数为 $\bar{x}_2 =$

$$\frac{99+99+100+100+100+102}{6} = 100,$$

甲、乙的数据的中位数均为 $\frac{100+100}{2} =$

100,

甲的数据的方差 $s_1^2 = \frac{1}{6} \sum_{i=1}^6 (x_i -$

$\bar{x}_1)^2 = \frac{7}{3}$, 乙的数据的方差 $s_2^2 =$

$$\frac{1}{6} \sum_{i=1}^6 (x_i - \bar{x}_2)^2 = 1,$$

甲的数据的极差为 $103-98=5$, 乙的数据

的极差为 $102-99=3$.

综上, 甲、乙的数据的平均数、中位数

相同, 甲的数据的方差大于乙的数据

的方差, 甲的数据的极差大于乙的数据

的极差.

故 A, B, D 错误, C 正确. 故选 C.

4. D 【解析】设数据 $2x_1+1, 2x_2+1, \dots,$

$2x_{2021}+1$ 平均数为 \bar{x}_0 , 方差为 s_0^2 , 则

$\bar{x}_0 = 2\bar{x}+1, s_0^2 = 4s^2$. 因为新数据的平均

数比方差大 3, 所以 $2\bar{x}+1 = 4s^2+3$, 可得

$$s^2 = \frac{1}{2}\bar{x} - \frac{1}{2},$$

则 $s^2 - \bar{x}^2 = \frac{1}{2}\bar{x} - \frac{1}{2} - \bar{x}^2 = -\left(\bar{x} - \frac{1}{4}\right)^2 -$

$\frac{7}{16}$. 由 $s^2 = \frac{1}{2}\bar{x} - \frac{1}{2} \geq 0$, 可得 $\bar{x} \geq 1$,

所以当 $\bar{x} = 1$ 时, $s^2 - \bar{x}^2$ 取最大值, 且

$s^2 - \bar{x}^2$ 的最大值为 -1 . 故选 D.

5. D 【解析】将数据从小到大依次排列

如下:

85, 87, 88, 89, 89, 90, 91, 91, 92, 93,

93, 93, 94, 96, 98,

$15 \times 60\% = 9, 15 \times 90\% = 13.5$,

故这组数据的 60 百分位数是 $\frac{1}{2} \times (92 +$

$93) = 92.5$, 这组数据的 90 百分位数

是 96. 故选 D.

6. ABC 【解析】由题中直方图可知此

次测试分数众数的估计值为 85, 故 A



正确;

因为不低于 80 分的学生的频率为

$$(0.030+0.015) \times 10 = 0.45,$$

所以该校高中生中随机抽取学生的人

$$\text{数为 } \frac{27}{0.45} = 60,$$

所以此次测试分数在 $[50, 60)$ 的学生

人数为 $60 \times 0.010 \times 10 = 6$, 故 B 正确;

因为 $(0.005 + 0.005 + 0.010 + 0.015 +$

$$0.020) \times 10 = 0.55,$$

所以随机抽取的学生测试分数的 55

百分位数约为 80, 故 C 正确;

由题中直方图可求 $m = 74$, $n =$

77.5 , $m < n$, 即平均数小于中位数, 即

平均数 m 在中位数 n 左侧, 故 D 错

误. 故选 ABC.

7. 17. 12 【解析】依题意可得总体平均

$$\text{数为 } \frac{22 \times 16 + 28 \times 18}{50} = 17.12.$$

8. 9. 2 【解析】由频率分布直方图可知,

样本中高中学生每天的平均学习时间的

60 百分位数在第 3 组,

设其为 x 小时,

$$\text{则 } 0.05 \times 2 + 0.10 \times 2 + 0.25(x - 8) =$$

$$0.6,$$

$$\text{解得 } x = 9.2,$$

所以估计该市高中学生每天的平均学

习时间的 60 百分位数为 9.2 小时.

9. 【解】(1) 由题设, 甲射击命中环数的

平均数为 $\bar{x}_1 = (7+8+6+8+6+5+9+10+$

$$7+4) \times \frac{1}{10} = 7,$$

乙射击命中环数的平均数为 $\bar{x}_2 = (9+$

$$5+7+8+7+6+8+6+7+7) \times \frac{1}{10} = 7.$$

(2) 甲射击命中环数的方差为 $s_1^2 =$

$$\frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x}_1)^2 = (0 + 1 + 1 + 1 +$$

$$1 + 4 + 4 + 9 + 0 + 9) \times \frac{1}{10} = 3,$$



乙射击命中环数的方差为 $s_2^2 =$

$$\frac{1}{10} \sum_{i=1}^{10} (x_i - \bar{x}_2)^2 = (4 + 4 + 0 + 1 + 0 + 1 + 1 + 1 + 0 + 0) \times \frac{1}{10} = 1.2.$$

由(1)知 $\bar{x}_1 = \bar{x}_2$, 而 $s_1^2 > s_2^2$,

所以选派乙去参赛更好.

10. 【解】(1) 由 $(0.005 + 0.010 + 0.020 + a + 0.025 + 0.010) \times 10 = 1$, 解得 $a = 0.030$.

(2) 因为 $(0.005 + 0.010 + 0.020) \times 10 = 0.35$, $(0.005 + 0.010 + 0.020 + 0.030) \times 10 = 0.65$,

所以样本数据的 62 百分位数在 $[70,$

$80)$ 内, 可得 $70 + \frac{0.62 - 0.35}{0.65 - 0.35} \times 10 = 79$,

所以样本数据的 62 百分位数为 79.

(3) 样本数据落在 $[50, 60)$ 内的个数为 $0.1 \times 100 = 10$, 落在 $[60, 70)$ 内的个数为 $0.2 \times 100 = 20$,

$$\begin{aligned} \bar{x} &= \frac{10}{10+20} \times 52 + \frac{20}{10+20} \times 64 = 60, s^2 = \\ &\frac{10}{10+20} \times [6 + (52-60)^2] + \frac{20}{10+20} \times [3 + (64-60)^2] = 36. \end{aligned}$$

11. C 【解析】若 5 个不同自然数从小到大依次为 a, b, c, d, e , 且中位数为 15, 平均数为 10, 则要使最大的数最大, 只需其他数最小, 即 $a = 0, b = 1, c = 15, d = 16$,

所以这 5 个自然数中最大数为 e , e 的最大值为 $10 \times 5 - 0 - 1 - 15 - 16 = 18$.