# 第九单元 统计与统计案例

## 基础课50 随机抽样与统计图表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **考点考向** | **课标要求** | **真题印证** | **考频热度** | **核心素养** |
| 随机抽样 | 理解 | 2023年新高考Ⅱ卷 | ★☆☆ | 数据分析  数学运算 |
| 统计图表 | 理解 | 2023年上海卷  2022年北京卷 | ★★★ | 数据分析  数学建模  数学运算 |
| 命题分析预测 | 从近几年高考的情况来看，命题热点是情境化试题.分层随机抽样出现的频率高，但题目难度不大.预计2025年高考命题情况变化不大，但应加强对创新问题的重视 | | | |

### 基础知识·诊断

#### 夯实基础

##### 一、简单随机抽样

1*.*定义:一般地,从*N*(*N*为正整数)个不同个体构成的总体中,逐个不放回地抽取*n*(1≤*n<N*)个个体组成样本,并且每次抽取时总体内的每个个体被抽到的可能性相等,这样的抽取方法叫作①简单随机抽样*.*

2*.*方法:②抽签法和③随机数法*.*

##### 二、分层随机抽样

1*.*定义:将总体按其属性特征分成互不交叉的若干类型(有时称作层),然后在每个类型中按照所占比例随机抽取一定的个体,这种抽样方法叫作④分层随机抽样*.*

2*.*适用范围:当总体由差异明显的几类个体构成,并且知道每一类个体在总体中所占的百分比时,通常采用分层随机抽样*.*

3*.*分层随机抽样的平均数:一般地,将样本*a*1,*a*2,…,*am*和样本*b*1,*b*2,…,*bn*合并成一个新样本,则*=*·*+*·*.*于是,当已知上述两层构成的样本中每层的平均数分别为和时,可得这个新样本的平均数为*+.*记*ω*1*=*,*ω*2*=*,则这个新样本的平均数为*ω*1*+ω*2,其中*ω*1,*ω*2称为权重*.*更一般地,设样本中不同层的平均数和相应的权重分别为,,…,和*ω*1,*ω*2,…,*ωn*,则这个样本的平均数为*ω*1*+ω*2*+*…*+ωn.*为了简化表示,引进求和符号,记作*ω*1*+ω*2*+*…*+ωn=ωi.*

4*.*分层随机抽样的方差:设样本中不同层的平均数分别为,,…,,方差分别为,,…,,相应的权重分别为*ω*1,*ω*2,…,*ωn*,则这个样本的方差为*s*2*=ωi*[*+*(*-*)2],其中为这个样本的平均数*.*

##### 三、作频率分布直方图的步骤

1*.*求⑤极差(即一组数据中最大值与最小值的差);

2*.*决定组距与组数;

3*.*将⑥数据分组;

4*.*列频率分布表;

5*.*画频率分布直方图*.*

#### 诊断自测

##### 题组1 走出误区

1. 判一判.（对的打“√”,错的打“×”）

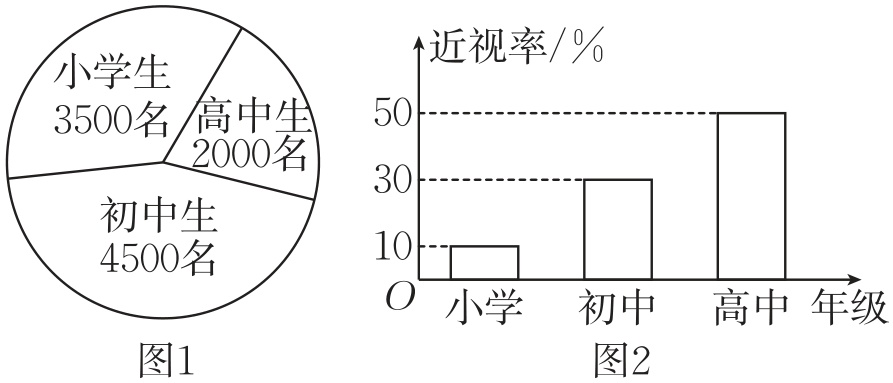
（1） 在简单随机抽样中，每个个体被抽到的机会不一样，与先后有关.( × )

（2） 抽签法和随机数法都是简单随机抽样.( √ )

（3） 在分层随机抽样中，每个个体被抽到的可能性与层数及分层有关.( × )

（4） 在频率分布直方图中，小长方形的面积越大，表示样本数据落在该区间的频率越大.( √ )

2. （易错题）已知某地区中小学的学生人数和近视情况分别如图1和图2所示，为了了解该地区中小学生的近视形成原因，用比例分配的分层随机抽样的方法抽取的学生进行调查，则样本容量和抽取的高中生近视人数分别为( B ).



A. 100，20 B. 200，20 C. 100，10 D. 200，10

**【易错点】**在比例分配的分层随机抽样中找不准比例标准致误，忽视两个图表之间的联系.

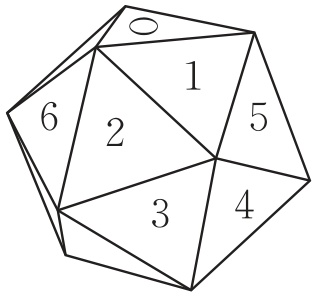
[解析]由题知样本容量为，其中高中生人数为，高中生近视人数为.故选.

##### 题组2 走进教材

3. （人教A版必修②P184·T3改编）已知高二年级有男生490人，女生510人，（张华）按男生、女生进行分层，通过分层随机抽样的方法，得到男生、女生的平均身高分别为和，若张华在各层中按比例分配样本，总样本量为100，则在这种情况下，高二年级全体学生的平均身高为165.4.（结果保留一位小数）

[解析]抽取的男生人数为，抽取的女生人数为.高二年级全体学生的平均身高估计为.

4. （人教A版必修②P177·T2改编）如图，这是一个正20面体（每个面都是正三角形），将20个面平分成10组，第1组标上0，第2组标上1， ，第10组标上9.三个正20面体分别涂上红、黄、蓝三种颜色，分别代表百位、十位、个位，若同时投掷可以产生一个三位数（百位为0的也看作三位数），则它产生的三位数的范围是.



[解析]三个正20面体分别涂上红、黄、蓝三种颜色，分别代表百位、十位、个位，同时投掷产生一个三位数（百位为0的也看作三位数），该三位数最大为999，最小为000，它是范围内的随机数.

##### 题组3 走向高考

5. [2023·新高考Ⅱ卷改编]某学校为了了解学生参加体育运动的情况，用比例分配的分层随机抽样做抽样调查，拟从初中部和高中部两层共抽取60名学生，已知该校初中部和高中部分别有400名和200名学生，则抽取的高中部学生有20名.

[解析]根据分层随机抽样的定义知高中部共抽取（名）.

### 考点聚焦·突破

#### 考点一 简单随机抽样［自主练透］

1. 下列抽样方法是简单随机抽样的是( D ).

A. 从平面直角坐标系中抽取5个点作为样本

B. 某饮料公司从仓库的1000瓶饮料中一次性抽取20瓶进行质量检查

C. 某连队从200名战士中，挑选出50名最优秀的战士参加抢险救灾

D. 从10部手机中不放回地随机抽取2部进行质量检验（假设10部手机已编好号，对编号随机抽取）

[解析]对于,平面直角坐标系中有无数个点，这与简单随机抽样要求总体中的个体数有限不相符，所以中的抽样方法不是简单随机抽样，故错误；对于,一次性抽取20瓶，不符合逐个抽取的特点，所以不是简单随机抽样，故错误；对于,挑选的50名战士是最优秀的，所以不符合简单随机抽样的等可能性，故中的抽样方法不是简单随机抽样，故错误；对于,易知中的抽样方法是简单随机抽样，故正确.故选.

2. 一个盒子中有若干白色的围棋子，为了估计其中围棋子的数目，小明将100颗黑色的围棋子放入其中，充分搅拌后随机抽出了20颗，其中有5颗黑色的围棋子，根据这些信息可以估计白色的围棋子有( B ).

A. 200颗 B. 300颗 C. 400颗 D. 500颗

[解析]设白色围棋子的数目为，则由已知可得，解得，即白色围棋子的数目大约有300颗.故选.

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

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

[解析]从随机数表第1行的第5列和第6列数字开始，从左到右依次选取两个数字，得65,72,08,02,63,14,07,02,43,69,97,28,01，98， ，去掉不在编号范围内的65，72，63，43，69，97，28，98，再去掉一个重复的02，得前5个个体的编号为08，02，14，07，01，故选出来的第5个个体的编号为01.



1.不放回简单随机抽样需满足：

（1）被抽取的总体的个体数有限；

（2）是逐个抽取；

（3）是不放回抽取；

（4）是等可能抽取.

2.简单随机抽样常用抽签法（适用于总体中个体数较少的情况）、随机数法（适用于总体中个体数较多的情况）.

#### 考点二 分层随机抽样［多维探究］

##### 分层随机抽样的比例分配角度1

典例1 [2024·安庆模拟]某工厂生产甲、乙、丙、丁四种不同型号的产品，产量分别为200,400,300,100件，为检验产品的质量，现用分层随机抽样的方法从以上所有的产品中抽取60件进行检验，则应从丙型号的产品中抽取18件.

[解析]因为样本容量，总体容量，所以抽取比例为，因此应从丙型号的产品中抽取（件）.



**分层随机抽样问题的类型及解题思路**

1.求某层应抽个体数量,按该层所占总体的比例计算.

2.已知某层个体数量求总体容量，可反之求解：根据分层随机抽样就是按比例抽样，列比例式进行计算.

3.分层随机抽样的计算应根据抽样比构造方程求解，其中，抽样比.

##### 分层随机抽样的样本均值与样本方差角度2

典例2 [2024·合肥模拟]有4万个大于70的两位数，从中随机抽取了3000个数，统计情况如表所示：

|  |  |  |  |
| --- | --- | --- | --- |
| 数据 |  |  |  |
| 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4万个数的平均数为85.23.

[解析]这3000个数的平均数为.于是由总体取值规律的估计得这4万个数的平均数为85.23.

##### 多维训练

1. 已知甲车间有工人54人，乙车间有工人42人，用分层随机抽样的方法从这两个车间中抽出一部分人组成方队，进行体能比赛，则甲车间和乙车间分别被抽取的人数是( A ).

A. 9，7 B. 15，1 C. 8，8 D. 12，4

[解析]由题意得甲车间被抽取的人数为，

乙车间被抽取的人数为.故选.

2. 某学校高一年级有300名男生，200名女生，通过分层随机抽样的方法调查高一年级学生的数学考试成绩，抽取总样本量为50，男生平均成绩为120分，女生平均成绩为110分，那么可以推测高一年级学生的数学平均成绩为( C ).

A. 110分 B. 115分 C. 116分 D. 120分

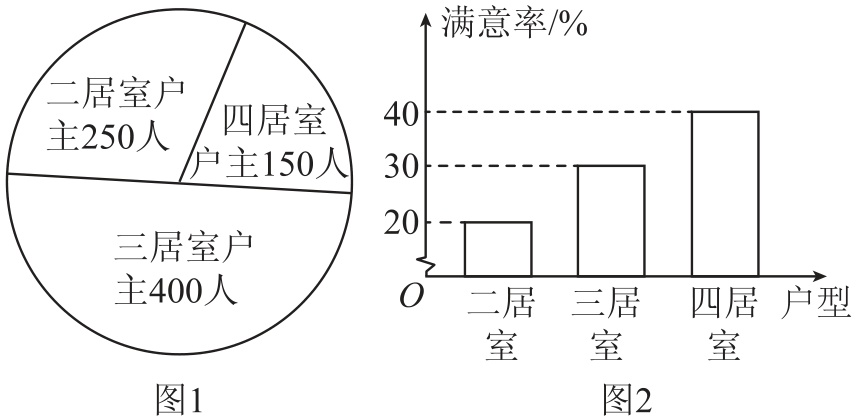
[解析]由题意，应抽取男生的人数为，

应抽取女生的人数为，所以推测高一年级学生的数学平均成绩为（分）.故选.

#### 考点三 统计图［多维探究］

##### 扇形图、条形图、折线图角度1

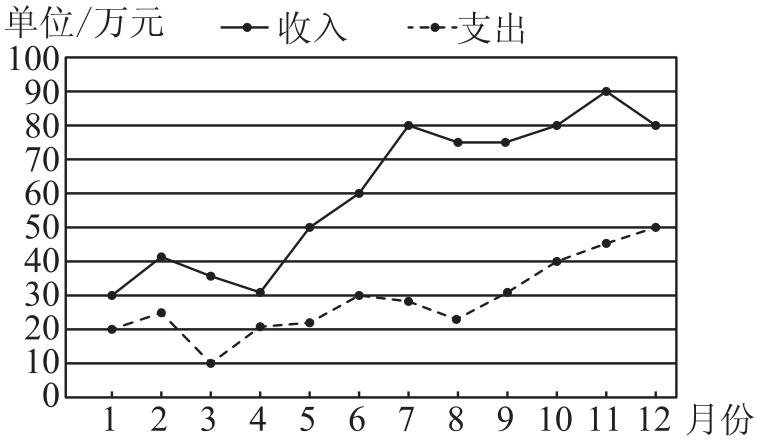
典例3（1） 已知某居民小区户主人数和户主对户型结构的满意率分别如图1和图2所示，为了了解该小区户主对户型结构的满意程度，用分层随机抽样的方法抽取的户主进行调查，则样本容量和抽取的户主中对四居室满意的人数分别为( A ).



A. 240，18 B. 200，20 C. 240，20 D. 200，18

[解析]样本量，抽取的户主对四居室满意的人数为.故选.

（2） [2024·湖北模拟]（多选题）某企业2022年12个月的收入与支出数据的折线图如图所示.已知利润收入-支出，根据该折线图，下列说法正确的是( ABC ).



A. 该企业2022年1月至6月的总利润低于2022年7月至12月的总利润

B. 该企业2022年1月至6月的平均收入低于2022年7月至12月的平均收入

C. 该企业2022年8月至12月的支出持续增长

D. 该企业2022年11月份的月利润最大

[解析]因为图中的实线与虚线的相对高度表示当月利润，所以由折线统计图可知，1月至6月的相对高度的总量要比7月至12月的相对高度的总量少，故正确；1月至6月的收入都普遍低于7月至12月的收入，故正确；2022年8月至12月的虚线是持续上升的，即支出持续增长，故正确；11月的相对高度比7月、8月都要小，故错误.故选.



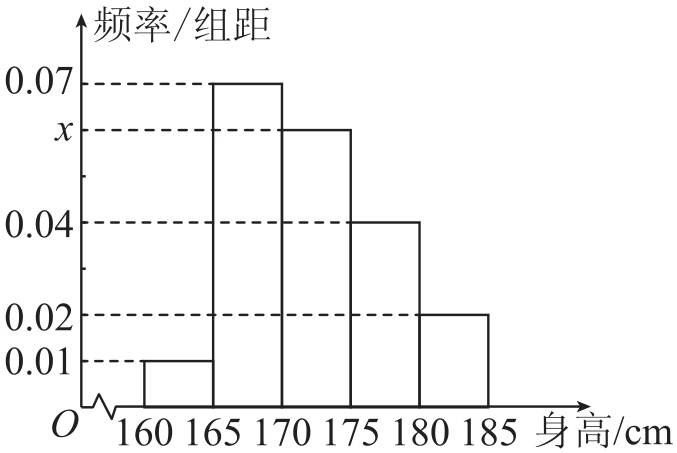
**统计图表问题的解决方法**

1.首先要准确地识图，即要明确统计图表中纵轴、横轴及折线、区域等所表示的意义，尤其注意数字变化的趋势等.

2.其次要准确地用图，会根据统计图表中的数字计算样本的数字特征，会用统计图表估计总体.

##### 频率分布直方图角度2

典例4 [2024·北京模拟]随机抽取100名学生，测得他们的身高（单位：），并按照区间，，分组，得到样本身高的频率分布直方图如图所示.



（1）求频率分布直方图中的值及身高在及以上的学生人数；

（2）将身高在区间，内的学生依次记为，，三个组，用分层随机抽样的方法从这三个组中抽取6人，求这三个组分别抽取的学生人数.

[解析]（1）由频率分布直方图可知，，解得，

身高在及以上的学生人数为.

（2）由题意可知，组人数为，组人数为，组人数为，

组抽取的人数为，

组抽取的人数为，

组抽取的人数为.



**频率分布直方图的相关结论**

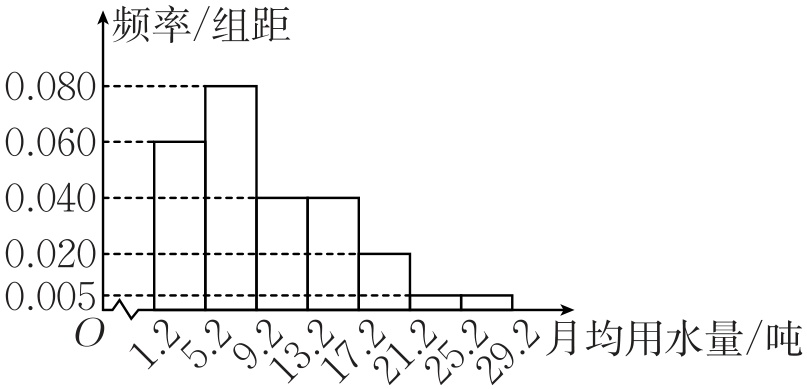
1.频率分布直方图中各小长方形的面积之和为1.

2.频率分布直方图中纵轴表示，故每组样本的频率为组距，即矩形的面积.

3.频率分布直方图中每组样本的频数为频率×总数.

##### 多维训练

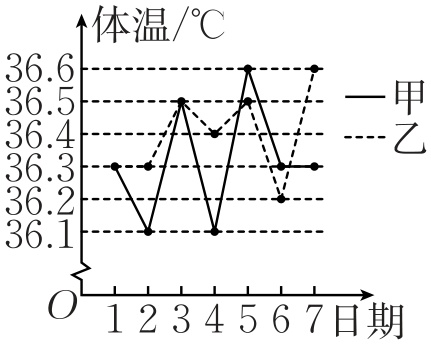
1. 为了减少水资源的浪费，某市计划对居民生活用水费用实施阶梯式水价制度.为了确定一个比较合理的标准，通过简单随机抽样，获得了100户居民的月均用水量数据（单位：吨），得到如图所示的频率分布直方图.估计该市居民月均用水量的中位数为( B ).



A. 8.25 B. 8.45 C. 8.65 D. 8.85

[解析]由频率分布直方图，得月均用水量在5.2吨以下的居民用户所占的比例为，月均用水量在9.2吨以下的居民用户所占的比例为，故中位数落在区间内.设样本的中位数为，则，所以，即样本的中位数为，由总体集中趋势的估计可知居民月均用水量的中位数为8.45.故选.

2. 某班级体温检测员对某一周内甲、乙两名同学的体温进行了统计，其结果如图所示，则下列结论不正确的是( C ).



A. 甲同学的体温的极差为

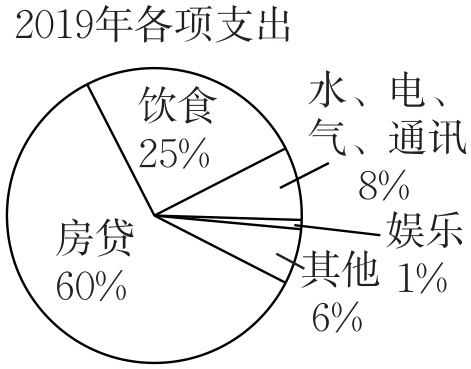
B. 甲同学的体温的众数为

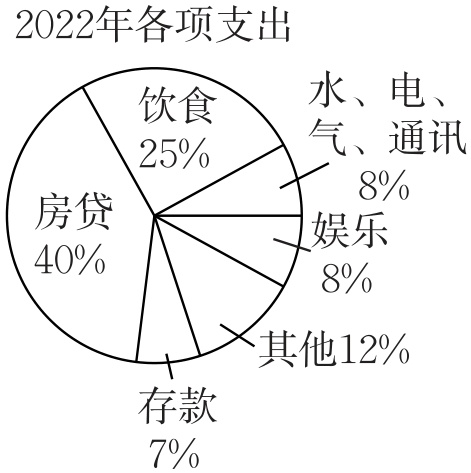
C. 乙同学的体温的中位数与平均数不相等

D. 乙同学的体温比甲同学的体温稳定

[解析]对于,甲同学的体温的极差为，故正确；对于,甲同学的体温从低到高依次为，，，，，，，故众数为，故正确；对于,乙同学的体温从低到高依次为，，，，，，，故中位数为，而平均数也是，故不正确；对于,从折线图上可以看出，乙同学的体温比甲同学的体温稳定，故正确.故选.

3. 小王于2018年底贷款购置了一套房子，根据家庭收入情况，小王选择了10年期每月还款数额相同的还贷方式，且截至2022年底，他没有再购买第二套房子.如图，这是2019年和2022年家庭收入用于各项支出的比例分配图，则下列结论正确的是( B ).





A. 小王一家2022年用于饮食的支出费用与2019年相同

B. 小王一家2022年用于其他方面的支出费用是2019年的3倍

C. 小王一家2022年的家庭收入比2019年增加了1倍

D. 小王一家2022年的房贷支出比2019年少

[解析]对于，小王一家2022年用于饮食的支出比例与2019年相同，但是由于2022年比2019年家庭收入多， 小王一家2022年用于饮食的支出费用比2019年多，故错误；对于，设2019年收入为， 相同的还款数额在2019年占各项支出的，在2022年占各项支出的，收入为， 小王一家2022年用于其他方面的支出费用为，小王一家2019年用于其他方面的支出费用为， 小王一家2022年用于其他方面的支出费用是2019年的3倍，故正确；对于，设2019年收入为，则2022年收入为，故错误；对于，小王一家2022年用于房贷的支出费用与2019年相同，故错误.故选.