## 14.2.2　分层抽样

学习目标　1.理解分层抽样的基本思想和适用情形.2.掌握分层抽样的实施步骤.3.了解简单随机抽样与分层抽样的区别和联系．

知识点一　分层抽样

1．分层抽样的定义

一般地，当总体由差异明显的几个部分组成时，为了使样本更客观地反映总体情况，我们常常将总体中的个体按不同的特点分成层次比较分明的几个部分，然后按各个部分在总体中所占的比实施抽样，这种抽样方法叫作分层抽样，所分成的各个部分称为“层”．

2．分层抽样的特点

(1)适用于总体由差异明显的几部分组成的情况．

(2)按比例确定每层抽取个体的个数．

(3)在每一层进行抽样时，采用简单随机抽样的方法．

(4)分层抽样能充分利用已掌握的信息，使样本具有良好的代表性．

(5)分层抽样也是等机会抽样，每个个体被抽到的可能性都是，而且在每层抽样时，可以根据个体情况采用不同的抽样方法．

知识点二　分层抽样的实施步骤

1．将总体按一定标准分层．

2．计算各层的个体数与总体的个体数的比．

3．按各层的个体数占总体的个体数的比确定各层应抽取的样本容量．

4．在每一层进行抽样(可用简单随机抽样)．

知识点三　两种抽样方法的比较

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 类别 | 特点 | 相互联系 | 适用范围 | 共同点 |
| 简单随机抽样 | 从总体中逐个抽取 |  | 总体中的个体数相对较少 | 抽样过程中每个个体被抽到的可能性相同 |
| 分层抽样 | 将总体分成几层，按各层的个体数之比抽取 | 各层抽样时，可以采用简单随机抽样 | 总体由差异明显的几部分组成 |

1．分层抽样是等可能抽样．(　√　)

2．分层抽样是按一定的比例从各层抽取个体组成样本的抽样．(　√　)

3．在分层抽样中，具体分多少层由具体情况而定．(　√　)

4．分层抽样中不含简单随机抽样．(　×　)

一、对分层抽样概念的理解

例1　①某班数学期中考试有14人在120分以上，35人在90～119分，7人不及格，现从中抽出8人研讨进一步改进教与学；②高一某班级春节聚会，要产生两位“幸运者”．就这两件事，合适的抽样方法分别为(　　)

A．分层抽样，简单随机抽样

B．简单随机抽样，分层抽样

C．简单随机抽样，简单随机抽样

D．分层抽样，分层抽样

答案　A

解析　①由于学生分成了差异比较大的几部分，应用分层抽样．②由于总体与样本容量较小，应用简单随机抽样．

反思感悟　判断抽样方法是分层抽样，主要是依据分层抽样的特点

(1)适用于总体由差异明显的几部分组成的情况．

(2)样本能更充分地反映总体的情况．

(3)等可能抽样，每个个体被抽到的可能性都相等．

跟踪训练1　在100个零件中，有一级品20个，二级品30个，三级品50个，从中抽取20个作为样本．

方法1：采用简单随机抽样的方法，将零件编号为00,01,02，…，99，用抽签法抽取20个．

方法2：采用分层抽样的方法，从一级品中随机抽取4个，从二级品中随机抽取6个，从三级品中随机抽取10个．

对于上述问题，下列说法正确的是\_\_\_\_\_\_\_\_．(填序号)

①不论采用哪种抽样方法，这100个零件中每一个零件被抽到的可能性都是；

②采用不同的方法，这100个零件中每一个零件被抽到的可能性各不相同；

③在上述两种抽样方法中，方法2抽到的样本比方法1抽到的样本更能反映总体特征．

答案　①③

解析　根据两种抽样的特点知，不论哪种抽样，总体中每个个体入样的可能性都相等，都是，故①正确，②错误；由于总体中有差异较明显的三个层(一级品、二级品和三级品)，故方法2抽到的样本更有代表性，③正确．故①③正确．

二、分层抽样的应用

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，职工年龄与这项指标有关，应该怎样抽取？

解　用分层抽样来抽取样本，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＝，则在不到35岁的职工中抽取125×＝25(人)；

在35岁至49岁的职工中抽取280×＝56(人)；

在50岁及50岁以上的职工中抽取95×＝19(人)．

(3)在各层分别按随机数表法抽取样本．

(4)汇总每层抽样，组成样本．

反思感悟　利用分层抽样抽取样本的操作步骤

(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．

(3)按各层的个体数占总体的比确定各层应抽取的样本容量．

(4)在每一层进行抽样(可用简单随机抽样)．

(5)最后将每一层抽取的样本汇总合成样本．

跟踪训练2　某市的3个区共有高中学生20 000人，且3个区的高中学生人数之比为2∶3∶5，现要从所有学生中抽取一个容量为200的样本，调查该市高中学生的视力情况，试写出抽样过程．

解　(1)由于该市高中学生的视力有差异，按3个区分成三层，用分层抽样来抽取样本．

(2)确定每层抽取个体的个数，在3个区分别抽取的学生人数之比也是2∶3∶5，所以抽取的学生人数分别是200×＝40；200×＝60；200×＝100.

(3)在各层分别按简单随机抽样抽取样本．

(4)综合每层抽样，组成容量为200的样本．

1．某单位有职工750人，其中青年职工350人，中年职工250人，老年职工150人，为了了解该单位职工的健康情况，用分层抽样的方法从中抽取样本．若样本中的青年职工为7人，则样本容量为(　　)

A．12 B．13 C．14 D．15

答案　D

解析　青年职工、中年职工、老年职工三层之比为7∶5∶3，所以样本容量为7÷＝15.

2．某校对全校1 200名男女学生进行健康调查，采用分层抽样法抽取一个容量为200的样本．已知女生抽了85人，则该校的男生人数为(　　)

A．670 B．680 C．690 D．700

答案　C

解析　男生人数占总人数的比等于抽到男生人数占样本容量的比，可得男生人数为1 200×＝690.

3．为了调查某省各城市PM2.5的值，按地域把36个城市分成甲、乙、丙三组，对应的城市数分别为6,12,18.若用分层抽样的方法抽取12个城市，则乙组中应抽取的城市数为\_\_\_\_\_\_\_\_．

答案　4

解析　乙组城市数占总城市数的比例为＝，样本容量为12，故乙组中应抽取的城市数为12×＝4.

4．某商场有四类食品，食品类别和种数见下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 类别 | 粮食类 | 植物油类 | 动物性食品类 | 果蔬类 |
| 种数 | 40 | 10 | 30 | 20 |

现从中抽取一个容量为20的样本进行食品安全检测．若采用分层抽样方法抽取样本，则抽取的植物油类与果蔬类食品种数之和为\_\_\_\_\_\_\_\_．

答案　6

解析　因为总体的个数为40＋10＋30＋20＝100，所以根据分层抽样的定义可知，抽取的植物油类食品种数为×20＝2，抽取的果蔬类食品种数为×20＝4，所以抽取的植物油类与果蔬类食品种数之和为2＋4＝6.

5．某企业三月中旬生产*A*，*B*，*C*三种产品共3 000件，根据分层抽样的结果，企业统计员制作了如下的表格：

|  |  |  |  |
| --- | --- | --- | --- |
| 产品类别 | *A* | *B* | *C* |
| 产品数量(件) |  | 1 300 |  |
| 样本数量(件) |  | 130 |  |

由于不小心，表格中*A*，*C*产品的有关数据已被污染看不清楚，统计员记得*A*产品的样本数量比*C*产品的样本数量多10，根据以上信息，可得*C*产品的数量是\_\_\_\_\_\_\_\_．

答案　800

解析　设*C*产品的样本数量为*n*，则*A*产品的样本数量为*n*＋10，由题意知＝，解得*n*＝80.

故*C*产品的数量为80÷＝800.

1．知识清单：

(1)分层抽样．

(2)分层抽样的应用．

2．方法归纳：数据分析．

3．常见误区：在分层抽样中，易忽略每个个体被抽到的可能性相等，与层数及分层无关而致错，每一层的抽样一般采用简单随机抽样．

1．下列实验中最适合用分层抽样法抽样的是(　　)

A．从一箱3 000个零件中抽取5个入样

B．从一箱3 000个零件中抽取600个入样

C．从一箱30个零件中抽取5个入样

D．从甲厂生产的100个零件和乙厂生产的200个零件中抽取6个入样

答案　D

解析　D中总体有明显差异，故用分层抽样．

2．将*A*，*B*，*C*三种性质的个体按1∶2∶4的比例进行分层抽样调查，若抽取的样本容量为21，则*A*，*B*，*C*三种性质的个体分别抽取的个数为(　　)

A．12,6,3 B．12,3,6

C．3,6,12 D．3,12,6

答案　C

解析　由分层抽样的概念，知*A*，*B*，*C*三种性质的个体应分别抽取的个数为21×＝3,21×＝6,21×＝12.

3．某中学有高中生3 500人，初中生1 500人，为了解学生的学习情况，用分层抽样的方法从该校学生中抽取一个容量为*n*的样本，已知从高中生中抽取70人，则*n*为(　　)

A．100 B．150

C．200 D．250

答案　A

解析　由题意得，＝，解得*n*＝100.

4．我国古代数学名著《九章算术》中有如下问题“今有北乡算八千七百五十八，西乡算七千二百三十六，南乡算八千三百五十六，凡三乡，发役三百七十八人，欲以算数多少出之，问各几何？”意思是：北乡有8 758人，西乡有7 236人，南乡有8 356人，现要按人数多少从三乡共征集378人，问从各乡征集多少人？在上述问题中，需从西乡征集的人数是(　　)

A．102 B．112 C．130 D．136

答案　B

解析　因为北乡有8 758人，西乡有7 236人，南乡有8 356人，现要按人数多少从三乡共征集378人，故需从西乡征集的人数是378×≈112.

5．某学校高一、高二、高三三个年级共有学生3 500人，其中高三学生人数是高一学生人数的两倍，高二学生人数比高一学生人数多300，现在按的抽样比用分层抽样的方法抽取样本，则应抽取的高一学生人数为(　　)

A．8 B．11 C．16 D．10

答案　A

解析　若设高三学生人数为*x*，则高一学生人数为，高二学生人数为＋300，所以有*x*＋＋＋300＝3 500，解得*x*＝1 600.故高一学生人数为800，因此应抽取的高一学生人数为＝8.

6．在1 000个球中有红球50个，从中抽取100个进行分析，如果用分层抽样的方法对球进行抽样，则应抽取红球的个数为\_\_\_\_\_\_\_\_．

答案　5

解析　设应抽红球*x*个，则＝，则*x*＝5.

7．某学校高一年级有*x*个学生，高二年级有*y*个学生，高三年级有*z*个学生．现采用分层抽样法抽取一个容量为45的样本，其中高一年级被抽取20人，高三年级被抽取10人，高二年级共有300人，则此学校共有高中学生\_\_\_\_\_\_人．

答案　900

解析　高二年级被抽取45－20－10＝15(人)，

被抽取的比例为＝，所以*x*＝400，*z*＝200.

所以此学校共有高中学生900人．

8．某橘子园有平地和山地共120亩，现在要估计平均亩产量，按一定的比例用分层抽样的方法共抽取10亩进行统计．如果所抽取的山地是平地的2倍多1亩，则这个橘子园的平地与山地的亩数分别为\_\_\_\_\_\_\_\_．

答案　36,84

解析　设所抽取的平地的亩数为*x*，则抽取的山地的亩数为2*x*＋1，∴*x*＋2*x*＋1＝10，得*x*＝3，∴这个橘子园的平地的亩数为120×＝36，山地的亩数为120－36＝84.

9．为了对某课题进行研究，分别从*A*，*B*，*C*三所高校中用分层抽样法抽取若干名教授组成研究小组，其中高校*A*有*m*名教授，高校*B*有72名教授，高校*C*有*n*名教授(其中0＜*m*≤72≤*n*)．

(1)若*A*，*B*两所高校中共抽取3名教授，*B*，*C*两所高校中共抽取5名教授，求*m*，*n*；

(2)若高校*B*中抽取的教授总数是高校*A*和*C*中抽取的教授总数的，求三所高校的教授的总人数．

解　(1)∵0＜*m*≤72≤*n*，*A*，*B*两所高校中共抽取3名教授，∴高校*B*中抽取2人，

∴高校*A*中抽取1人，高校*C*中抽取3人，

∴＝＝，解得*m*＝36，*n*＝108.

(2)∵高校*B*中抽取的教授总数是高校*A*和*C*中抽取的教授总数的，

∴(*m*＋*n*)＝72，解得*m*＋*n*＝108，

∴三所高校的教授的总人数为*m*＋*n*＋72＝180.

10．某中学举行了为期3天的新世纪体育运动会，同时进行全校精神文明擂台赛．为了解这次活动在全校师生中产生的影响，分别在全校500名教职员工、3 000名初中生、4 000名高中生中做问卷调查，如果要在所有答卷中抽出120份用于评估．

(1)应如何抽取才能得到比较客观的评价结论？

(2)要从3 000份初中生的答卷中抽取一个容量为48的样本，如果采用简单随机抽样，应如何操作？

解　(1)由于这次活动对教职员工、初中生和高中生产生的影响不会相同，所以应当采取分层抽样的方法进行抽样．

因为样本容量为120，总体个数为500＋3 000＋4 000＝7 500，则抽样比为＝，

所以500×＝8,3 000×＝48,4 000×＝64，所以在教职员工、初中生、高中生中抽取的个体数分别是8,48,64.

分层抽样的步骤是

①分层：将全校师生分为教职员工、初中生、高中生，共三层．

②确定每层抽取个体的个数：在教职员工、初中生、高中生中抽取的个体数分别是8,48,64.

③各层分别按简单随机抽样的方法抽取样本．

④综合每层抽样，组成样本．

这样便完成了整个抽样过程，就能得到比较客观的评价结论．

(2)由于简单随机抽样有两种方法：抽签法和随机数表法．如果用抽签法，要作3 000个号签，费时费力，因此采用随机数表法抽取样本，步骤是

①编号：将3 000份答卷都编上号码：0001,0002,0003，…，3000.

②在随机数表上随机选取一个起始位置．

③规定读数方向：向右连续取数字，以4个数为一组，如果读取的4位数大于3000，则去掉，如果遇到相同号码则只取一个，这样一直到取满48个号码为止．

11．某校共有学生2 000名，各年级男、女生人数如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 一年级 | 二年级 | 三年级 |
| 女生 | 373 | 380 | *y* |
| 男生 | 377 | 370 | *z* |

现用分层抽样的方法在全校抽取64名学生，则应在三年级抽取的学生人数为(　　)

A．24 B．18 C．16 D．12

答案　C

解析　一年级的学生人数为373＋377＝750，二年级的学生人数为380＋370＝750，于是三年级的学生人数为2 000－750－750＝500，那么三年级应抽取的人数为500×＝16.

12．某校做了一次关于“感恩父母”的问卷调查，从8～10岁，11～12岁，13～14岁，15～16岁四个年龄段回收的问卷依次为：120份，180份，240份，*x*份．因调查需要，从回收的问卷中按年龄段分层抽取容量为300的样本，其中在11～12岁学生问卷中抽取60份，则在15～16岁学生中抽取的问卷份数为(　　)

A．60 B．80 C．120 D．180

答案　C

解析　11～12岁回收180份，其中在11～12岁学生问卷中抽取60份，抽样比为，因为分层抽取的样本容量为300，故回收问卷总数为＝900(份)，故*x*＝900－120－180－240＝360(份)，360×＝120(份)．

13．某机构对某镇的学生的身体素质状况按年级段进行分层抽样调查，得到了如下表所示的数据，则＝\_\_\_\_\_\_.

|  |  |  |  |
| --- | --- | --- | --- |
| 年级段 | 小学 | 初中 | 高中 |
| 总人数 | 800 | *x* | *y* |
| 样本中人数 | 16 | 15 | *z* |

答案　37 500

解析　由分层抽样的特点，得＝＝，即*x*＝750，＝50，则＝37 500.

14．某高中针对学生发展要求，开设了富有地方特色的“泥塑”与“剪纸”两个社团，已知报名参加这两个社团的学生共有800人，按照要求每人只能参加一个社团，各年级参加社团的人数情况如下表：

|  |  |  |  |
| --- | --- | --- | --- |
|  | 高一年级 | 高二年级 | 高三年级 |
| 泥塑 | *a* | *b* | *c* |
| 剪纸 | *x* | *y* | *z* |

其中*x*∶*y*∶*z*＝5∶3∶2，且“泥塑”社团的人数占两个社团总人数的，为了了解学生对两个社团活动的满意程度，从中抽取一个50人的样本进行调查，则从高二年级“剪纸”社团的学生中应抽取\_\_\_\_\_\_\_\_人．

答案　6

解析　因为“泥塑”社团的人数占总人数的，故“剪纸”社团的人数占总人数的，所以“剪纸”社团的人数为800×＝320.因为“剪纸”社团中高二年级人数比例为＝＝，所以“剪纸”社团中高二年级人数为320×＝96.由题意知，抽样比为＝，所以从高二年级“剪纸”社团中抽取的人数为96×＝6.

15．已知某地区中小学生人数和近视情况分别如图①和图②所示．为了解该地区中小学生的近视形成原因，用分层抽样的方法抽取2%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(　　)

A．200,20 B．100,20

C．200,10 D．100,10

答案　A

解析　该地区中小学生总人数为3 500＋2 000＋4 500＝10 000，

则样本容量为10 000×2%＝200，其中抽取的高中生近视人数为2 000×2%×50%＝20.

16．某单位最近组织了一次健身活动，活动分为登山组和游泳组，且每个职工至多参加其中一组，在参加活动的职工中，青年人占42.5%，中年人占47.5%，老年人占10%.登山组的职工占参加活动总人数的，且该组中青年人占50%，中年人占40%，老年人占10%，为了了解各组不同年龄层次的职工对本次活动满意程度，现用分层抽样的方法从参加活动的全体职工中抽取一个容量为200的样本，试确定：

(1)游泳组中，青年人、中年人、老年人分别所占的比例；

(2)游泳组中，青年人、中年人、老年人分别应抽取的人数．

解　(1)设登山组人数为*x*，游泳组中，青年人、中年人、老年人所占比例分别为*a*，*b*，*c*，

则有＝47.5%，＝10%，

解得*b*＝50%，*c*＝10%，

故*a*＝100%－50%－10%＝40%.

即游泳组中，青年人、中年人、老年人所占比例分别为40%,50%,10%.

(2)游泳组中，

抽取的青年人人数为200××40%＝60；

抽取的中年人人数为200××50%＝75；

抽取的老年人人数为200××10%＝15.

即游泳组中，青年人、中年人、老年人分别应抽取的人数为60,75,15.