**中国矿业大学2016～2017学年第二学期**

**《抽样调查》试卷（A）卷**

**考试时间：120分钟 考试方式：闭卷**

学院 班级 姓名 学号

|  |  |  |  |
| --- | --- | --- | --- |
| 题号 | 一 | 二 | 总分 |
| 得分 |  |  |  |
| 阅卷人 |  | | |

*注：邱松强为命题教师。*

一、简答题：（本题共6小题，每小题5分，满分30分）

**评分说明：以下提供的是答案要点，视学生答题情况，根据其解答与以下要点切合程度酌情给分。**

1、阐述估计量方差、偏差和均方误差的含义.

估计量方差衡量估计值与待估参数的差异状况，其表达式为

偏差是指按照某一抽样方案反复进行抽样，估计值的数学期望与待估参数之间的偏差. 偏差的表达式为

均方误差指所有可能的估计值和待估参数之间离差平方的均值，其表达式为

2、简述（不放回）简单随机抽样的抽取规则和估计量  方差 .

(1) 按随机原则取样；(2) 每个抽样单元被抽中的概率都是已知或事先确定的，或者可以事先计算出来;(3) 每个抽样单元被抽中的概率都相等.

其中, f 为抽样比，为有限总体的方差.

3、简述分层抽样的定义和特点.

分层抽样是将抽样单元按某种特征或某种规则划分为不同的层，然后从不同底层中独立、随机地抽取样本，将各层样本结合起来，对总体的目标量进行估计.

分层抽样的优点：可以有效地提高估计的精度；在一定条件下可为组织实施调查提供方便.

4、整群抽样的优缺点是什么？

整群抽样的特点：抽取样本时只需要群的抽样框而不必要求具有所有基本单位的抽样框; 群通常由那些地理位置邻近的或隶属于同一系统的单元构成，因此调查地点比较集中，方便实施调查.

整群抽样的缺点是精度较差.

5、简述直线等距抽样和圆形等距抽样的规则. ()

直线等距抽样：总体中的N个单元已按某种确定顺序编号为1,2,…N. 先从头k个单元编号中随机抽出一个单元编号，然后每隔k个单元编号抽出一个单元编号， 直到抽出n个单元编号为止。

圆形等距抽样：假设总体总体中的N个单元已按某种确定顺序编号为 1,2,…N，将这些编号看成首尾相接的一个环，并从1到N中按简单随机抽样方式抽取一个单元编号作为随机起点，然后每隔 k抽取一个单元编号，直到抽满n个单元为止

6、 为什么多阶抽样常在大型抽样调查中被采用？

多阶段抽样具有整群抽样的优点，保证了样本相对集中，从而节约了调查费用；不需要包含所有低阶段抽样单元的抽样框；同时由于实行了再抽样，使调查单元在更大范围内展开。因此在大规模抽样调查中，多阶段抽样是经常采用的方法.

二、综合题：（本题共7小题，满分70分）(可能用到的分位点数 )

1、(5分) 设总体有5个单元（1、2、3、4、5），按不放回简单随机抽样的方式抽取2个单元，列出所有可能的样本.

解：如下标

5分

2、(10 分) 从某个100的总体中抽出一个大小为10的简单随机样本，试估计总体平均水平并给出置信度为95%的区间估计。

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 序号 *i* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 4 | 5 | 2 | 3 | 4 | 6 | 6 | 15 | 0 | 9 |

解：

3分

则抽样方差为

6分

置信区间为

10分

3、(10 分) 某居委会辖有三个居民新村，居委会欲对居民购买彩票的情况进行调查. 调查者考虑以新村分层，在每个新村随机抽取了10个居民并调查每户最近一个月购买彩票所花费的金额（元），下表是每个新村及调查的情况：

|  |  |  |  |
| --- | --- | --- | --- |
| 新村 | 居民户数 | 均值 | 方差 |
| 1 | 256 | 11.2 | 94.4 |
| 2 | 420 | 25.5 | 302.5 |
| 3 | 168 | 20 | 355.56 |

试估计该小区居民户购买彩票的平均支出，并给出估计的标准差.

解：

2分

4分

其方差为

8分

标准差为

10分

4、(10 分) 为了便于管理，将某林区划分为386 个小区域. 现采用简单随机抽样方法，从中抽取10个小区域，测量树的高度，得到如下资料

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 区域编号 | 树木株数 | 平均高度 | 区域编号 | 树木株数 | 平均高度 |
| 1 | 42 | 6.2 | 6 | 58 | 6.9 |
| 2 | 51 | 5.8 | 7 | 43 | 4.3 |
| 3 | 49 | 5.7 | 8 | 59 | 5.2 |
| 4 | 55 | 4.9 | 9 | 48 | 5.7 |
| 5 | 47 | 5.2 | 10 | 41 | 6.1 |

试估计整个林区树的平均高度及95%的置信区间.

解：



3分

7分

置信区间为

10分

5、(15 分) 对某个总体进行调查，事先规定了每个单元被抽中的概率 ，且已知 与规模 成正比. 今利用代码法抽取一个 的PPS 样本。

表：利用代码法抽样

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *i* |  |  | 累计 | 代码 |
| 1 | 0.104 | 104 | 104 | 0~104 |
| 2 | 0.192 | 192 | 296 | 105~296 |
| 3 | 0.138 | 138 | 434 | 297~434 |
| 4 | 0.062 | 62 | 496 | 435~496 |
| 5 | 0.052 | 52 | 548 | 497~548 |
| 6 | 0.147 | 147 | 695 | 549~695 |
| 7 | 0.089 | 89 | 784 | 696~784 |
| 8 | 0.038 | 38 | 822 | 785~822 |
| 9 | 0.057 | 57 | 879 | 823~879 |
| 10 | 0.121 | 121 | 1000 | 880~1000 |

（1）将上表补充完整

（2）假设抽样系统随机产生三个代码 300，487，991. 求相应的被抽中的单元.

(3) 经调查抽得的三个单元的值分别是 320， 120，290. 试估计总体总量并计算估计的方差.

解：(1) 见上表 5分

(2) 由上表，300，487，991 对应的抽样单元为3，4，10 10分

(3) 总体总量的估计值为

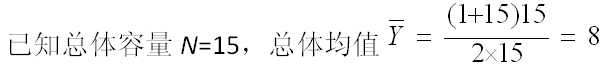
 12分

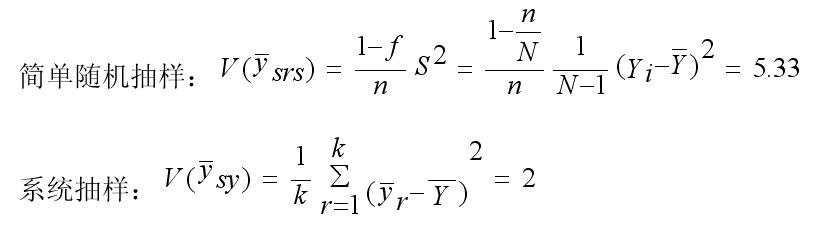
总量估计的标准差为

15分

6.(10 分) 假设总体 , 相应指标值排列顺序为 1，2，3，4，5，6，7，8，9，10，11，12，13，14，15. 考虑 的直线系统抽样，试计算系统抽样的实际方差，并于样本量相同的简单随机抽样进行比较.

解：

 6分

 10分

7.(10 分)某地调查农作物的总产量，该地共有大小不等的土地140块，总面积为460亩，采用简单随机抽样调查了10块土地，数据如下：

（1）当估计量是简单估计量时，估计该地平均产量及其标准差

（2）当估计量是比估计量时，估计该地的平均产量及其标准差

（3）上述结果的比较及其分析。

解: (1)  

;

4分

(2)

8分

(3) 通过上述计算，比估计量的标准差比简单随机抽样的标准差小，因而更为精确。 10分