**中国矿业大学2019-2020学年第一学期**

**《抽样调查》试卷（B）卷**

**考试时间：100分钟 考试方式：闭卷**

学院 班级 姓名 学号

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题号 | 一 | 二 | 三 | 四 | 总分 |
| 得分 |  |  |  |  |  |

1. 简答（本大题共3小题，每小题10分，共30分）
2. 抽样理论和数理统计中关于样本均值的定义和性质有哪些不同？

解析：抽样理论和数理统计中关于样本均值的定义和性质的不同

|  |  |  |
| --- | --- | --- |
|  | 抽样理论 | 概率统计 |
| 定义 |  |  |
| 性质 | 1. 期望 2. 方差 | 1. 期望      1. 方差 |

1. 请列举出四种基本的概率抽样方法，并简要说明各种抽样方法的思想。

简单随机抽样：从抽样框内随机地，逐个抽取 n 个单元作为样本，在每次抽选中，所有未入样的待选单元的入样概率都相等。

分层抽样： 将样本单元按某种特征划分为不同的层，然后从不同的层中独立、随机地抽取样本，将各层的样本结合起来，对总体的目标量进行评估。

整群抽样：将总体中若干个基本单元合并为组，这样的组称为群。抽样时直接抽取群，然后对选中的群全部实施调查。

多阶段调查：采取类似整群抽样的方法，首先抽取群，但不是调查群内的所有基本单元，而是再进一步抽样，从选中的群中抽取若干个基本单元进行调查。

（系统抽样：将总体中的所有单元按一定顺序排列，在规定的范围内随机抽取一个单元作为初始单元，然后按事先规定号的规则确定其他样本单元）

评分说明：答对一个得3分，两个得5分，三个得8分，四个得10分。

1. 简述PPS抽样的代码法和拉系里法。

代码法：在PPS抽样中，赋予每个单元与相等的代码数，将代码数累加得到M0，每次抽样都产生一个[1 M0]之间的随机数，设为m，则代码所对应的单元被抽中。如此进行n次，就构成了PPS抽样的样本。如果Mi不是整数，则乘以某个倍数，使M0Zi 为整数，每个单元赋予与M0Zi 相等的代码数，然后进行代码法抽样。 ……4分

拉希里法：令M\*=max{Mi},即所有Mi中最大值，每次抽样都产生一个[1 N]之间的随机数i和[1 M\*]之间随机数m, 如果Mi大于等于m，则第i个单元被抽中；否则，重抽一组（i，m）。反复进行以上操作，直至抽满n 个样本为止。

二、选择题(从下列各题四个备选答案中选出一个正确答案，并将其代号写在括号内。答案选错或未选者，该题不得分。每小题3分,共15分)

1. 概率抽样与非概率抽样的根本区别是：　　　（　　　　）

A．是否能确保总体中的每个单位都有完全相同的概率被抽中；

B．是否能确保总体中的每个单位都有事先已知或可以计算的非零概率被抽中；

C．是否能减少调查性误差；

D．是否能计算和控制抽样误差．

2．在要求的精度水平下，不考虑其他因素的影响，若简单随机抽样所需要的样本量为400，分层随机抽样的设计效应deff=0.7，则若想达到相同的精度，分层随机抽样所需要的样本量为：　 （　　　　）

A．240；　B. 280； C．320；　　D. 360．

3．当为某一特定常数时,回归估计可以看成差估计,此时该常数值为( )

A.  B.  C.  D. 

4．在系统抽样中，若总体采用无关标志排列，则系统抽样的效果等同于： （ ）

A.分层抽样；　　　 　B.简单随机抽样；

C.不等概率抽样；　　　　　D.整群抽样．

5．在估计总体均值时，无回答偏差的大小取决于：　　　　　　　　　　（　　　　）

A.回答层均值与无回答层均值之间的差异；

B.无回答者绝对数量的多少；

C.无回答率的高低；

D.无回答率的高低和回答层均值与无回答层均值之间的差异．

答： 1 B，2 B，3 C，4 B，5 D

三、计算题：（本题共4小题，每题10分，满分40分）(可能用到的分位点数 )

1、 为了合理调配电力资源，某市欲了解50 000 户居民的日用电量，从中简单随机抽取了300 户进行调查，现得到其日用电量平均值值为 (千瓦时)，. 试估计该市居民日常平均用电量的 95% 置信区间. 如果希望相对误差不超过 10%，则样本量至少应为多少？

答：

  5分

故而置信区间为

 7分

要使相对误差不超过10%，即

解得 ，即最少需要862个样本. 10分

说明：公式正确而计算错误可以酌情给分.

1. 某市有甲、乙两个地区，现要进行家庭收入的调查，令，已知甲地区共有20000户居民，乙地区共有50000户居民；甲地居民和乙地居民年收入标准差估计分别为；同时对甲地和乙地每户的平均抽样费用之比为2：3，请分别计算出在甲地和乙地进行比例分配、一般最优分配（考虑费用因素）以及内曼分配（不考虑费用因素）的样本量。

解：比例 143,357；

一般最优：190,310；

内曼：167,334

3、某开发区利用电话调查（RDD）对区内居民消费冷冻食品情况进行调查. 按局号分层，按每个局号拥有的电话数比例分配样本量（各层抽样比可以忽略）. 调查后各层样本用户购买冷冻食品支出的中间结果－表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 局号 | 层权（%） | 样本量 | 样本平均(元) | 样本标准差 |
| 1 | 14.5 | 29 | 73 | 90 |
| 2 | 19.0 | 38 | 89 | 142 |
| 3 | 23.5 | 47 | 83 | 97 |
| 4 | 27.0 | 54 | 68 | 92 |
| 5 | 16.0 | 32 | 44 | 55 |

试估计该开发区居民户购买冷冻食品的平均支出，并给出估计95%置信区间.

解：分层样本，总体均值的估计为



将上表数值代入得 样本均值的方差为



由于各层抽样比可以忽略，我们得 .

这个估计的95% 置信区间为



4、邮局欲估计每个家庭的平均订报份数，该辖区共有4000户，划分为400个群，每群10户，现随机抽取4个群，取得资料如下表所示：

|  |  |  |
| --- | --- | --- |
| 群 | 各户订报数 |  |
| 1 | 1，2，1，3，3，2，1，4，1，1 | 19 |
| 2 | 1，3，2，2，3，1，4，1，1，2 | 20 |
| 3 | 2，1，1，1，1，3，2，1，3，1 | 16 |
| 4 | 1，1，3，2，1，5，1，2，3，1 | 20 |

试估计平均每户家庭订报份数及总的订报份数，以及估计量的方差。

解：由题意得到，，，

故（份）

（份）

（份）











于是由以上的计算结果得到平均每户的订报份数为1.875，估计量方差为0.00391875。该辖区总的订阅份数为7500，估计量方差为62700。

1. 证明题（15分）证明简单随机抽样总体均值简单估计量的方差为：

