**中国矿业大学2019-2020学年第一学期**

**《抽样调查》试卷（A）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 | 五 | 总分 |
| 得分 |  |  |  |  |  |  |

1. 名词解释（本大题共5小题，每小题4分，共20分）
2. 总体 2.样本 3.抽样误差 4.估计量 5. PPS抽样

答：

1. 总体：有目标总体与调查总体之分。目标总体由符合研究目的的所有具有相同属性或特征的单位所组成，是抽样推断的目标所在。调查总体由已被人们所掌握、能对其加以排序或编号的单位所组成，是真正从中抽取样本的总体。
2. 样本：是总体的一部分，是从抽样总体中按一定方法和程序抽取的部分调查单位的集合，也称为子样。
3. 抽样误差：由于抽样的非全面性和随机性所引起的偶然性的代表性误差。
4. 估计量：是以样本指标为基础构造的、用以估计总体指标的规则或形式。优良估计量有三个标准：无偏性、一致性和有效性。
5. PPS抽样：按照总体的单元的规模大小来确定单元每次入样的概率，这种不等概抽样称为放回的与规模大小成比例的概率抽样。

二、填空题（本大题共15空，每空1分，共15分；请在每小题的空格中填上正确答案。错填、不填均无分。）

1.影响抽样误差大小的因素有 、 、 、 。

2.常用的概率抽样方式有 、 、 、 、 。

3.样本中所包含的调查单位数称为 。

4.简单随机抽样中，样本均值的方差等于 。

5.回归估计的应用条件要求 的总体总量或总体均值应该是已知的。

6.分层抽样中样本分配方法有 、 、 。

答1. 总体内在差异；样本容量；抽样方法、方式；估计量。

2. 简单随机抽样、分层抽样、系统抽样、整群抽样、多阶段抽样

3. 样本容量

4.

5. 辅助变量

6. 比例分配、最优分配、内曼分配

三、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2分,共10分)

1、抽样调查的根本功能是( )

A.获取样本资料 B. 计算样本资料

C.推断总体数量特征 D. 节约费用

2、当为某一特定常数时,回归估计可以看成比率估计,此时该常数值为( )

A.  B.  C.  D. 

3、抽样设计中最好的方案是( )

A. 抽样误差最小 B. 调查单位最少

C. 调查费用最省 D. 一定误差下使得费用最省

4、抽样误差的大小与下列哪个因素无关（ ）

A．样本容量 B．抽样方式、方法

C．概率保证程度 D．估计量

5、下面式子中错误的是 ( )

A.  B. 

C.  D. 

答： 1．C 2．A 3．D 4、C 5．D

四、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、 研究某小区家庭用于文化方面（报刊、电视、网络、书籍等）的支出，N=200，现抽取一个容量为20的样本，调查结果列于下表：

|  |  |  |  |
| --- | --- | --- | --- |
| 编号 | 文化支出 | 编号 | 文化支出 |
| 1 | 200 | 11 | 150 |
| 2 | 150 | 12 | 160 |
| 3 | 170 | 13 | 180 |
| 4 | 150 | 14 | 130 |
| 5 | 160 | 15 | 100 |
| 6 | 130 | 16 | 180 |
| 7 | 140 | 17 | 100 |
| 8 | 100 | 18 | 180 |
| 9 | 110 | 19 | 170 |
| 10 | 220 | 20 | 120 |

估计该小区平均的文化支出,并给出置信水平95%的置信区间。

解：由已知得： 

根据表中数据计算得： (3分)

（3分）

 （2分）

因此该小区平均文化支出的95%置信区间为：，即是：[135.85 164.15]。 （2分）

故估计该小区平均的文化支出=150,置信水平95%的置信区间为[135.85 ,164.15]。

2、某地区对本地100家化肥生产企业的尿素产量进行调查，已知去年的总产量为2345吨，抽取10个企业调查今年的产量，得到，这些企业去年的平均产量为。试估计今年该地区化肥总产量。

解:由题可知，, (2分)

则该地区化肥产量均值的比率估计量（4分）

该地区化肥产量总值Y的比率估计量（2分）

所以，今年该地区化肥总产量的估计值为2626.4吨。

3、调查某地区居民的奶制品年消费支出，以居民户为抽样单元，根据经济及收入水平将居民户划分为4层，每层按简单随机抽样抽取10户，调查后各层样本户的奶制品年消费支出的中间结果如下表所示：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 层号（） | 层权（%） | 样本量 | 样本平均值（元） | 第层样本方差 |
| 1 | 0.0702 | 10 | 39.5 | 1623.7222 |
| 2 | 0.1404 | 10 | 105.5 | 2166.6667 |
| 3 | 0.2632 | 10 | 165.0 | 8205.5556 |
| 4 | 0.5263 | 10 | 23.0 | 193.333 |

请估计该地区居民奶制品的平均消费总支出及其95%的置信区间。

解:根据表中的数据可得； (3分)

估计量方差的无偏估计为，（3分）

进而有。 （2分）

的95%的置信区间为，其中=1.96，经计算可得



 （2分）

因此，可以有95%的把握认为该地区居民奶制品的年消费总支出在57.046-89.19元之间。

4、 某市为了了解职工收入情况，从该市的630个企业中随机抽取了5个企业，在中选的企业中对职工在进行随机抽样，有关数据如下：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 企业号 |  |  | （元） |  |
| 1 | 520 | 10 | 328 | 400.056 |
| 2 | 108 | 10 | 400 | 301.134 |
| 3 | 1400 | 20 | 310 | 1303.158 |
| 4 | 1200 | 20 | 370 | 1205.786 |
| 5 | 9000 | 100 | 420 | 4200.000 |

其中，为企业职工数，为样本量；为样本均值，为样本方差。

试估计该市职工平均收入及标准差。

解：已知：N = 630，n = 5，

估计该市职工的平均收入为：

 (5分)

估计该市职工平均收入的方差及标准差为：



（5分）

因此，估计该市职工平均收入为398元，标准差为26.71669元

1. （15分）证明 对于两阶段抽样，有  式中， ,分别表示在固定初级单元时对第二阶段抽样求均值和方差； ，分别表示对第一阶段抽样求均值和方差。

证明：对于均值公式可以理解为对所有可能样本的平均，可以分两步进行。在给定的一个样本容量为n的初级单元样本中，对所有两阶段抽样可能的样本估计量进行平均，然后再对所有一阶段抽样可能的样本估计量进行平均。

记由对两边求，得

