**中国矿业大学2018-2019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 | 总分 |
| 得分 |  |  |  |
| 阅卷人 |  | | |

1. 简答题：（本题共5小题，每小题8分，满分40分）
2. 抽样误差和非抽样误差各自产生的原因是什么？我们该如何避免或者降低这两类误差?

答：抽样误差是由于用样本数据对总体特征进行估计所引起的代表性误差。由于每次只是抽取一个样本，而这个样本中包含哪些单元是随机的。不同的样本由于包含的单元不同，得到的估计值自然不同。各个估计值与总体特征之间不可避免的存在差距，由此产生了抽样误差。 …4分

非抽样误差的来源比较复杂，主要由有抽样框未能不重不漏的包含所有抽样单元导致的抽样框误差，调查测量不准确引致的测量误差，此外还有无回答误差、粗大误差等。非抽样误差不仅在抽样调查中，在全面调查中也是存在的。减小非抽样误差的方法主要是严格调查程序、规范调查步骤、加强人员的培训和管理，合理地设计问题和答卷，改进测量方法和工具。 ……4分

2、阐述估计量方差、偏差和均方误差的含义。

答：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 ……3分

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 ……3分

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分

3、简述分层随机抽样的具体抽样方法及特点，并比较它与整群抽样方法的共同点与区别。

答：分层随机抽样：事先按照一定的规则对总体进行分层。如果每层中的抽样都是独立地按照简单随机抽样进行的，那么这样的分层抽样就成为分层随机抽样。有三个必要条件：（1）每层都抽样；（2）各层独立进行抽样；（3）各层的抽样都是简单随机抽样。 ……2分

整群抽样：先将总体划分为若干群，然后以群为抽样单元，从总体中随机抽取一部分群，再对入选的群内的所有单元进行调查。 ……2分

联系：都是先将总体进行划分，然后抽取一部分单元，都是采用简单随机抽样。 ……2分

区别：分层抽样是各层都抽样，而整群抽样只抽取部分群；分层时层间差异大，层内差异小；

分群时群内差异大，群间差异小。 ……2分

4、简述二重抽样的概念及它与两阶段抽样的区别和联系。

答：二重抽样，也称二相抽样或两相抽样，是指在抽样时候分两步抽取样本，每一步抽取一个样本。一般情况下，先从总体N中抽取一个较大的样本，称为第一重样本，对其进行调查以获得总体的某些辅助信息，为下一步的抽样估计提供条件。然后进行第二重抽样，第二重抽样所抽的样本相对较小，但第二重抽样调查才是主调查。 ……4分

区别：首先，两阶段抽样是先从总体N个单元中抽出的样本单元，但并不对这个样本单元中的所有小单元都进行调查，而是在其中再抽出若干个二级单元进行调查；二重抽样则不同，要对第一重样本进行调查以获取总体的某些辅助信息，并且要利用这些辅助信息进行排序、分层、抽样或估计。

其次，两阶段抽样的第一阶段抽样单位和第二阶段抽样单位往往是不同的；而二重抽样的第二重样本则往往是第一重样本的子样本，两次抽样的单位是相同的。

联系：他们都进行两次抽样。

5、简述直线等距抽样和圆形等距抽样的规则 ()。

 答：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。

……4分

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。 ……4分

二、计算题：（本题共4小题，满分60分）(可能用到的分位点数 )

1、(15 分) 某地区500个乡为了调查今年粮食总产量的估计，采用简单随机抽样技术调查了50个乡今年的粮食产量，得到（吨），。试据此估计该地区今年的粮食总产量，并给出置信水平为95%的置信区间。

解：由题意可知 C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps6CFA.tmp.png

则该地区的粮食总产量为

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps6CFB.tmp.png

其方差为

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps6CFC.tmp.png ………（10分）

所以该居民区总用水量95%的置信区间为：

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps6CFD.tmp.png, 即

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps6D0D.tmp.png …… （5分）

2、(15 分) 对某地区171 980户居民家庭收入进行调查，以居民户为抽样单位，根据城镇和乡村将居民划为2层，每层按简单随机抽样抽取800户，经整理得如下数据：

|  |  |  |  |
| --- | --- | --- | --- |
| 层 |  |  |  |
| 城镇 | 23560 | 15180 | 2972 |
| 乡村 | 148420 | 9856 | 2546 |

试根据此估计：

（1）居民平均收入及其95%的置信区间。

（2）若是将上述样本量1600按比例分配方法和内曼分配方法来重新进行分配，则各层样本量分别应为多少？

解：（1）由题中相关数据资料：C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E48.tmp.png

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E49.tmp.png

该地区居民平均收入的90%的置信区间为：

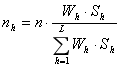
C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E4A.tmp.png ，即 C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E4B.tmp.png ……………（7分）

（2）按比例分配：

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E4C.tmp.png

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E4D.tmp.png ……………（4分）

按奈曼分配：



由表中资料：C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E4F.tmp.png

由上可得根据奈曼分配，各层所需样本容量为：

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E60.tmp.png

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wps4E61.tmp.png ……………（4分）

3、(15 分) 某居委会欲了解居民健身活动情况，如果已知该居委会有500名居民，居住在10个单元中。现先抽取4个单元，然后在样本单元中分别抽出若干居民，两个阶段的抽样都是简单随机抽样，调查得到的样本居民每天用于健身锻炼的时间结果如下（以10分钟为1个单位）：

|  |  |  |  |
| --- | --- | --- | --- |
| 单元 | 居民人数 | 样本量 | 健身锻炼时间 |
| 1 | 32 | 4 | 4, 2, 3, 6 |
| 2 | 45 | 5 | 2, 2, 4, 3, 6 |
| 3 | 36 | 4 | 3, 2, 5, 8 |
| 4 | 54 | 6 | 4, 3, 6, 2, 4, 6 |

试估计居民平均每天用于锻炼的时间，并给出估计的标准差。

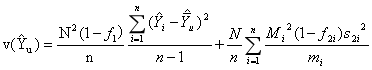
1. 简单估计量；
2. 比率估计量；
3. 对两种估计方法及估计结果进行评价。

解：（1）简单估计

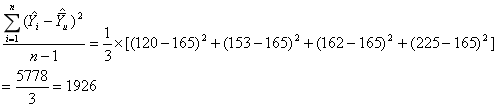
C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF55C.tmp.png=C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF55D.tmp.png

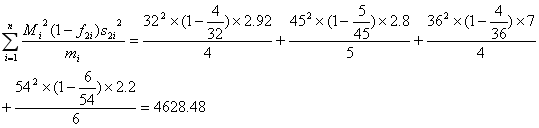
=1650，

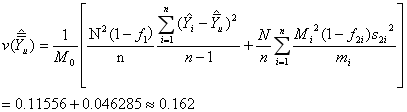
则C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF55E.tmp.png，又C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF55F.tmp.png，

所以

分别计算

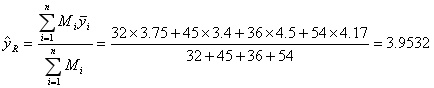


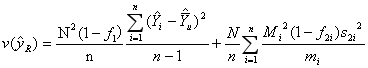


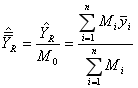
所以，

所以标准差C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF574.tmp.png

(2) 比率估计





其中

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF578.tmp.png

C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF589.tmp.png

（3）C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF58A.tmp.png简单估计标准差C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF58B.tmp.png，比率估计标准差C:\Users\songzhu\AppData\Local\Temp\ksohtml\wpsF58C.tmp.png

比率估计更好

1. (15 分) 某公司欲了解职工上班交通所需要的时间，该公司共有5个部门，根据每个部分的人数采用PPS抽样抽出2个部门，并在2个部门中采用简单随机抽样分别抽出5名职工，调查结果如下：

|  |  |  |
| --- | --- | --- |
| 部门 | 职工人数 | 时间（分钟） |
| 1 | 20 | 1. 10， 20， 30， 40 |
| 2 | 35 | 1. 30， 20， 60， 30 |

试估计该公司职工上班交通平均所需要时间，并计算估计得标准差。

解：





则该公司职工上班交通平均所需的时间为34分钟，

估计的标准差为6分钟。