

简述题A

1. 有人认为“抽样调查除了调查误差以外，还有抽样误差，因此抽样调查不如全面调查准确”，请对此加以评价
2. 什么是不完备抽样框？列举各种可能出现的情况以及对抽样推断的影响
3. 什么是抽样框？简述抽样框的类型和作用，以及良好抽样框的标志
4. 介绍简单随机抽样，包括特点，作用和应用场合

简单随机抽样(simple random sampling)也称纯随机抽样，是从抽样框内的N个抽样单元中随机地、一个一个地抽取n个单元作为样本，在每次抽选中，所有未入样的待选单元入选样本的概率相等，这n个被抽中的单元就构成了简单随机样本。简单随机样本也可以一次从总体（抽样框）中同时抽出，这时全部可能样本中的每一个样本被抽中的概率也需要相等。抽样的随机性通过抽样的随机化程序体现。

简单随机抽样是一种最基本的抽样方法，是其他抽样方法的基础，是任何其他概率抽样的核心内容。这种方法的突出特点是简单直观，在抽样框完整时，可以直接从中抽选样本，由于抽选的概率相同，用样本统计量对目标量进行估计及计算抽样误差都比较方便。在实际应用中也有一些局限：

- (1)它要求包含所有总体单元的名单作为抽样框，当N很大时，构造这样的抽样框并不容易；
- (2)根据这种方法抽出的单元很分散，给实施调查增加了困难；
- (3)这种方法没有利用其他辅助信息以提高估计的效率。

所以在规模较大的调查中，很少直接采用简单随机抽样，一般是把这种方法和其他抽样方法结合在一起使用。例如：分层抽样在每层内部均采用简单随机抽样，整群抽样是以群为单位进行简单随机抽样。此外，在许多日常场合，例如相当普遍的“摇号”等方式用于决定稀有物品或机会的归属等也是简单随机抽样的方式。

5. 说明总体方差和估计量方差各自的作用，以及它们之间的相互关系
6. 什么是样本量？试对影响样本量的因素进行分析

从总体中抽取若干单元的集合的数量叫样本量，用n表示。决定样本量大小的因素从定性分析角度有如下几点：

- (1) 调查经费：调查经费总数有限的，样本量不能无限大。
- (2) 总体方差：如果总体方差越大，在其他条件相同的情况下，需要的样本量也大。总体方差大，总体单元的差异大，抽样是随机的，有可能抽到大的，有可能抽到小的，样本的结果波动就比较大，误差就会比较大。所以为了保证其他的相同，样本量就需要比较大。
- (3) 允许误差范围：delta大小，在其他条件都相同条件，允许误差不同，样本量也不同
- (4) 置信度：通常取95%，不同置信度t值也不一样

7. 试分析简单随机抽样条件下影响估计量精度的因素
8. 试述抽样调查中产生偏差的原因以及如何对待这些偏差

偏差是指按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离差，表达式为

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

偏差是偏于某个方向的系统性误差，大多并不随样本量增大而减小。偏差产生原因有两种情况：

1. 估计量本身是有偏的，这时估计量的数学期望与总体参数不一致；
2. 非抽样误差因素的影响。

按照来源性质不同，非抽样误差分为三类：抽样框误差，无回答误差，计量误差。对于抽样框误差，可以采取三类补救措施TODO

在没有偏差的情况下，用样本统计量对目标量进行估计，要求估计量的方差越小越好；如果存在偏差，就需要把估计量方差和偏差综合起来考虑均方误差MSE。由于偏差属于系统误差，因而在抽样调查中应该努力避免。但是，也有一些估计量是有偏的，然后因偏差小带来的估计量方差也比较小，从而使均方误差比较小，这时选择这些有偏的估计量并不是一件坏事。

9. 根据样本量计算公式分析影响样本量因素

样本量计算步骤如下

- 所需要的精度
- 找出样本量与精度之间的关系
- 估计所需的数值，求解 n
- 如超出预算，调整精度值重新计算

费用限制

样本量费用函数为 $C = c_0 + c_1 n$, 总费用 C , 固定费用 c_0 , 单位费用 c_1 一定的情况下，样本量 n 就确定

$$n = \frac{C - c_0}{c_1}$$

精度要求

对精度的要求可以用绝对误差 d 和相对误差 r 表示，样本量足够大时，可以正态分布近似

$$d = t \sqrt{V(\hat{\theta})}$$

对于简单随机抽样，抽样方差只与总体特征和样本量 n 相关，按上面的步骤，均值估计时样本量为

1. 所需要的精度为 $P(|\hat{\theta} - \theta| \leq d) = 1 - \alpha$
2. 找出样本量和精度的公式 $V(\bar{y}) = (1 - f) \frac{S^2}{n}$

$$n_0 = \frac{t^2 S^2}{d^2} = \frac{t^2 S^2}{r^2 \bar{Y}^2}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

10. 试述概率抽样，非概率抽样的各自特点、作用和局限

概率抽样的特点：

1. 按随机原则抽取样本
2. 单元被选入样本的概率已知
3. 可以计算抽样误差

概率抽样用于统计推断，侧重点的分析

除了概率抽样以外，其他都叫做非概率抽样。主要特征是抽取样本时并不依据随机原则，单元入样概率不知，无法计算抽样误差，不具备统计推断意义。常见非概率抽样方式有

1. 方便选择
2. 目的选择
3. 自愿样本

4. 配额选择 非概率抽样的优点是操作简单，不需要抽样框，经济快速。

应用背景：

探索问题的研究，定性问题的研究。

11. 什么是分层随机抽样？试述分层抽样的特点和应用条件 在抽样之前，先将抽样单元按某种特征或某种规则划分为不同的层，然后从不同的层中分别独立地抽取样本，将各层的样本结合起来，对总体的目标量进行估计。这种抽样就是分层抽样。

- 分层抽样的抽样效率较高，也就是说分层抽样的估计精度较高。这是因为分层抽样估计量的方差只和层内方差有关，和层间方差无关。
- 分层抽样不仅能对总体指标进行推算，而且能对各层指标进行推算。
- 层内抽样方法可以不同，而且便于抽样工作的组织

应用条件：

- 估计：层内单元具有相同性质，通常按调查对象的不同类型进行划分。
- 目标总体应具有这样的结构：尽可能使层内单元的指标值相近，层间单元的差异尽可能大，从而达到提高抽样估计精度的目的。
- 估计和精度：既按类型、又按层内单元指标值相近的原则进行多重分层，同时达到实现估计类值以及提高估计精度的目的。
- 实施：抽样组织实施的方便，通常按行政管理机构设置进行分层。

12. 对分层抽样可以提高估计效率的原理进行论述

13. 在分层随机抽样下影响估计量精度的因素有哪些

分层抽样的估计量精度，即分层抽样的估计量方差，只是与层内方差有关，与层间方差无关。理由如下：

影响估计精度的因素有样本量、总体大小和总体方差。在其他因素不变的情况下，总体方差越大，估计的精度越差；反之，估计的精度越高。总体方差是客观存在且无法改变的。但对总体单元进行分层，即分成若干个子总体，根据方差分析原理，有：总平方和(SST) = 层间平方和(SSB) + 层内平方和(SSW) 根据上式，如果在子总体（层）内单元之间比较相似，每一个子总体的方差较小而各层子总体之间的方差较大，那么这样只需在各层子总体中抽取少量样本单元，就能很好地代表各层子总体的特征。所以，分层抽样的估计量方差只是与层内方差有关，与层间方差无关。在分层抽样时，各层子总体由于单元之间的差异小，且小于整体总体方差，而抽样精度与此成正比，所以分层抽样可以提高参数估计的精度。事实上，只要准确选择分层依据的指标，则分层抽样的精度一般不仅要高于简单随机抽样，也高于其他抽样方式。以总体均值估计量为例，分层随机抽样总体均值估计量的样本方差有：

$$V(\bar{y}_{st}) =$$

由此可以看出，分层抽样的估计量精度，即分层抽样的估计量方差与层权、样本单元数、每层抽样比、层内方差有关，与层间方差无关。其中层内方差是主要影响因素。

14. 简述比率估计方法、特点和应用条件

简述题B

1. 试对分层抽样中的联合比率估计和分别比率估计方法进行比较
主要是比较两者的方差

分别比估计

$$V(\bar{y}_{Rs}) =$$

联合比估计

$$V(\bar{y}_{cs}) =$$

区别在于R不一样。

1. 样本量足够，分别比估计较好
2. 各层样本量小，联合比估计较好
3. 若各层比估计量 r_i 相似，二种方法差别不大。

2. 试述在什么情况下需要采用不等概率抽样，并举例说明

3. 简述整群抽样的特点和应用背景

特点：

- 抽样框编制得以简化
- 实施调查便利：群内各个单元在地域上是接近的，实施比较便利
- 估计效率较低：因为群内各个单元存在同质性，但实际上调查一个单元可以得到整个群的情况，样本量很大，实际上效率较低
- 对某些特殊结构的总体有比较好的效果

4. 从统计学原理角度，对分层抽样的分层原则和整群抽样的分群原则论述

分层抽样的分层原则是：层内方差尽可能小，层间方差尽可能大；

而整群抽样的分群原则是：群内差异尽可能大，群间差异尽可能小。

理由如下：

分层抽样：

整群抽样：

由于整群抽样对入选群中的所有单元都进行调查，因此影响整群抽样误差主要受到群间方差的影响。为了提高其估计的精度，划分群的原则应当是使得群内方差尽可能大，不同群之间的差异尽可能小，也即分群时使得同一群内各单元的差异尽可能大，不同群之间的差异尽可能小。目的是能保证每个群对总体都具有足够好的代表性。极端情况下，当各群基本单元的分布完全相同，即群间不存在任何差异时，那么只需抽取一个群进行调查就能充分满足抽样估计精度的要求，整群抽样的效率就很高。这和分层抽样的原则恰好相反。

由此看来，整群抽样和分层抽样是针对不同总体结构而提出的两种不同抽样方式。对一些复杂结构的总体，可以把两种抽样方式结合起来，以发挥各自的特点。

5. 简述多阶段抽样的应用背景和特点

6. 试述按有关标志排列时系统抽样的效率

按无关标志排列时，按简单随机抽样方差计算

$$V(\bar{y}_{sy}) = \frac{(1-f)S^2}{n}$$

按有关标志排列时，误差

$$V(\bar{y}_{sy}) = \frac{S^2}{n} [1 + (n-1)\rho]$$

若样本单元分布均匀, $\rho < 0$

若样本单元分布随机, $\rho = 0$, 趋近于srs

若样本单元分布聚集, $\rho > 0$

总体单元按有关标志排列是指总体单元的排列顺序与所研究的内容有关系。一般常见的是将总体单元按标志值从小到大（或从大到小）顺序排列，或按某个与其线性相关的辅助变量的大小顺序排列。按有关标志排列方式的抽样称为有序系统抽样。对于按有关标志排列的线性趋势总体，系统抽样（等距抽样）的效率优于简单随机抽样，劣于按比例分配分层抽样。

- 与简单随机抽样相比 当总体单元按线性排列时，能使标志值高低不同的单位，均有可能选入样本，从而提高样本的代表性，增大系统样本内方差。在总体方差一定时，系统样本内方差越大，系统抽样估计量的方差越小，精度越高，故系统抽样估计量的方差小于简单随机抽样估计量的方差。
- 与按比例分层抽样相比 系统抽样可以看作在各层抽一个样本单元，若起始位置 r 偏小，则样本均值偏小；若起始位置 r 偏大，则样本均值偏大。在总体单元按有关标志排列后，总体各层单元分布集聚，样本单元密切程度越高，同一系统样本内对层均值离差的相关系数 $\rho_{wst} > 0$ 。但是分层抽样在各层的样本位置是随机的，因此不同位置对指标值的影响相互抵消，所以系统抽样估计量方差大于按比例分层抽样估计量的方差。

7. 自加权设计的方法

8. 当总体排列有周期性波动时，采用等距抽样应注意哪些问题，以及处理方法

9. 试述二重抽样的作用、特点和实践中需要注意的问题

作用：

- 进行样本筛选：
- 进行事后分层，抽选好的样本，使得样本结构与总体结构一致
- 获得辅助信息的估计，提高估计效率
- 用于对无回答的调整：有时调查拒访率较高，二重抽样可以降低无回答的偏倚影响

实践中需要注意：

- 费用效益的经济情况是在二重抽样中十分值得注意的问题。在抽取第一重大样本时需要增加一定的费用，只有精度的得益大于所增加的费用，采用二重抽样才是合算的。
- 此外，样本量的最优分配也是在二重抽样中值得注意的问题。二重抽样中需要两次抽取样本量，每次抽样都需要一定的费用。第一重抽样的样本量越大，对辅助信息的估计就越精确，但当总费用固定时，第一重抽样若占用过多的费用必然影响第二重抽样，使得第二重抽样样本量过小，从而影响估计量的精度。

10. 比较二阶段抽样和二重抽样的异同

共同点：

- 两种抽样方法都是经过两个阶段抽样最终样本单元确定
- 抽样误差具有相同特征：形式上都是两个误差的相加 不同点：
- 目标不同：二阶段抽样的第一阶段没法抽样，所以进行二阶段抽样，构造抽样框比较方便，样本相对集中，调查便于实施；而二重抽样可以直接抽样，他的目的或者是为了获得辅助信息，或者是为了进行分层，或者是为了从无回答层中获取无回答信息，总之有他特定的目标，和二阶段抽样目的不同。

- 抽样框不同：二阶段抽样的第一阶段抽样是使用psu作为抽样框，然后再中选的群中再采用最终单元的抽样框；而二重阶段第一阶段的抽样就是整个最终单元的抽样框。所以两种抽样方法抽样框构造是有差异的
- 由于数据不同，抽样框的来源资料也不同，所以最后误差的计算不完全相同

11. 简述辅助信息在抽样调查中的作用

12. 试述如何利用辅助信息改进估计量

13. 试述随机化回答技术作用、特点，谈谈你对随机化回答技术的理解
是针对敏感问题或者高度隐私问题的调查，采用的一种特殊方法。

特点：

- 随机回答的方式：问题是随机的
- 随即回答目的是对受访者起保护作用
- 同时可以对目标参数进行推断

使用随机化模型应当注意的问题

1. 高素质的调查人员：这种调查有特殊性，所以对调查人员素质要求比较高，要对受访者情况很好的了解。
2. 有效的说明工作：向受访者解释如何实施随机化回答技术，为什么这么做，可以对他的真实情况起到保密的作用。要让受访者配合调查，只有受访者了解了，相信了，才能配合你。所以有效的说明工作非常重要。调查人员素质越高，解释说明工作越明白。
3. 注意选择无关的非敏感性问题：在西蒙斯模型中看到，卡片A是敏感性问题，卡片B是非敏感性问题。如何选择好的非敏感性问题，总体结果是已知的，而问题的设计又要符合调查的受访人的场景，使他很好的放松他的心情，也是在问题的设计中很重要的一环。

14. 简述捕获再捕获方法，以及该方法的特征和应用场合

捕获再捕获(capture-recapture methods)是一种估计总体单元数的方法。该方法的实施流程为：先捕获数量为 n_1 的样本，并在样本上进行标记，然后将这些样本单元放回总体，在已标记的样本单元充分融入总体后，进行第二次捕获，这时又会得到样本量为 n 的样本，第二次捕获的才是正式样本，考察样本中有标记的单元个数，然后对总体单元数进行推断。捕获再捕获法要满足几个基本假定：

1. 总体是封闭的，即总体单元数在前后再次抽样中不变。
2. 两次抽样都是简单随机抽样，即每个单元在前后两次抽样中入样的机会是均等的。
3. 前后再次抽样是相互独立的，第一次抽样中被标记的单元放回总体后要充分混合以保证第二次抽样中
4. 被标记单元的标记不会丢失，保证第二次抽样中样本的可识别性。

捕获再捕获的方法最先应用于野生动物学中，用于监测鸟类、鱼类和害虫的数量。近年来，常被应用在

15. 比较分层抽样的分层原则和整群抽样的分群原则