

第七章 抽样调查

一、本章重点

1. 抽样调查也叫做抽样推断或参数估计，必须坚持随机抽样的原则。它是一种非全面调查，其意义在于对总体的推断上，存在可控制性误差。是一种灵活快捷的调查方式。

2. 抽样调查有全及总体与样本总体之区分。样本容量小于 30 时一般称为小样本。对于抽样调查来讲全及总体的指标叫做母体参数，是唯一确定的未知的量，样本指标是根据样本总体各单位标志值计算的综合性指标，是样本的一个函数，是一个随机变量，抽样调查就是要用样本指标去估计相应的总体指标。样本可能数目与样本容量有关也与抽样的方法有关。抽样方法可以分为考虑顺序的抽样与不考虑顺序的抽样；重复抽样与不重复抽样。

3. 大数定律、正态分布理论、中心极限定理是抽样调查的数理基础。正态分布的密度函数有两个重要的参数 $(\bar{x}; \sigma)$ 。它有对称性、非负性等特点。中心极限定理证明了所有样本指标的平均数等于总体指标如 $E(\bar{x}) = \bar{X}$ 。推出了样本分布的标准差为：

$$\mu_x = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}。$$

4. 抽样推断在逻辑上使用的是归纳推理的方法、在方法上使用的是概率估计的方法、存在着一定误差。无偏性、一致性和有效性是抽样估计的优良标准。

抽样调查既有登记性误差，也有代表性误差，抽样误差是一个随机变量，而抽样的平均误差是一个确定的值。抽样误差受总体标志值的差异程度、样本容量、抽样方法、抽样组织形式的影响。

在重复抽样下抽样的平均误差与总体标志值的差异程度成正比，与样本容量的平方根

成反比即 $\mu_x = \frac{\sigma}{\sqrt{n}}$ ，不重复抽样的抽样平均误差仅与重复抽样的平均误差相差一个修正因

子即 $\mu_x = \frac{\sigma}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$ 。在通常情况下总体的方差是未知的，一般要用样本的方差来代替。

把抽样调查中允许的误差范围称作抽样的极限误差 Δ_x 或 Δ_p 。 $\Delta = t\mu$ ，用抽样的平均误差来度量抽样的极限误差。把抽样估计的把握程度称为抽样估计的置信度。抽样的极限误差越大，抽样估计的置信度也越大。抽样估计又可区分为点估计和区间估计。按估计的指标不同又可分为总体平均数的估计、总体成数的估计和总体方差的估计。

二、难点释疑

1.要区分样本可能数目与必要抽样数目。样本可能数目是指从总体 N 中抽取一个样本容量为 n 的子样最多有多少种抽法，一般用 M 表示。而必要抽样数目则是为了使抽样误差控制在一定的范围内，至少应抽取多少个单位作样本，是样本容量 (n) 的另一种表现形式。

2.大数定律、正态分布理论、中心极限定理都是假定从 N 中抽取一个样本容量为 n 的子样，把所有的样本都抽到 (有 M 种抽法) 之后进行验证的，在实际工作中不可能办到。只能用样本的相应指标去推测总体的相应指标。这些理论只是为了验证抽样推断的科学性。

3.在实际工作中往往是以重复抽样的方法确定必要抽样数目，以不重复抽样的方法来抽取调查单位，进行计算估计，而又用重复抽样的误差公式来计算误差，一方面是计算公式简单，另一方面，这样计算的误差比实际存在的误差大，便于提高抽样推断的可靠性。

4.在抽样调查中总体的方差是未知的，一般都用样本的方差来代替。对于抽样成数来说，当 $p=0.5$ 时，抽样成数的方差取极大值 0.25。

三、练习题

(一) 填空题

1.抽样调查是建立在 () 上的一种科学的调查方法。它是从研究对象中 () 部分单位进行调查，以其达到对 () 的认识，并可对 () 进行计算和控制。

2.样本中所包含的总体单位数称为 ()，通常记为 n,当 n () 时称为大样本。

3.在抽样调查中总体指标是 ()，而样本指标则是随机的。

4.对确定的总体来说，可能的样本数目与 () 有关，也与 () 有关。

5.一般地说，用抽样指标估计总体指标应该有三个要求。这三个要求是 ()、()、()。

6.常用的抽样方法有（）抽样和（）抽样；（）抽样和（）抽样。

7.常用的抽样组织形式有（）、（）、（）、（）、（）。

8.抽样平均误差就是抽样平均数（或抽样成数）的（）。它反映抽样平均数（或抽样成数）与总体平均数（或总体成数）的（）。

9.抽样误差是一个（），抽样平均误差是一个（）。前者是在不考虑（）条件下，抽样估计值与被估计的真实总体参数之差；后者反映抽样平均数（成数）与总体平均数（成数）的（）。

10.影响抽样误差大小的因素有：（）、（）、抽样方法和（）。

11.对于同一种抽样组织形式来讲，重复抽样的平均误差（）不重复抽样的平均误差。当（）很小时，可以用重复抽样的平均误差公式来近似计算不重复抽样的平均误差。

12.在其它条件不变的情况下，极限误差增大，则概率度（），可靠程度（）。

13.在缺少总体的方差时，可用（）代替来计算抽样误差。

14.抽样极限误差是指抽样指标与总体指标之间抽样误差的（）。抽样极限误差通常以（）为标准单位来衡量。

（二）名词解释

1.随机性原则 2.抽样误差 3.样本 4.抽样平均误差

5.抽样极限误差 6.区间估计 7.重复抽样 8.不重复抽样

（三）判断题

1.抽样调查就是凭主观意识，从总体中抽取部分单位进行调查。（）

2.抽样法不仅是对被研究总体有关指标进行科学推算的方法，而且是搜集统计资料的方法。（）

3.抽样调查中的随机误差是可以避免的。（）

4.所有可能的样本平均数的平均数等于总体的平均数。（）

5.抽样误差是不可避免的，但人们可以调整总体的方差的大小来控制抽样误差的大小。（）

6.缩小抽样的极限误差便会降低推断的把握程度。（）

7.由于抽样调查中既有登记性误差又有代表性误差，所以只有登记性误差的全面调查准确性高。（）

8.重复抽样和不重复抽样是抽取样本的两种方法，各抽样组织方式都可以使用这两种

方法。()

(四) 单项选择题

1. 随机抽样的基本要求是严格遵守 ()

- A、准确性原则 B、随机性原则
- C、代表性原则 D、可靠性原则

2. 抽样调查的主要目的是 ()

- A、广泛运用数学方法 B、计算和控制抽样误差
- C、修正普查资料 D、用样本指标来推算总体指标

3. 抽样调查中 ()

- A、既有登记性误差，也有代表性误差
- B、只有登记性误差，没有代表性误差
- C、没有登记性误差，只有代表性误差
- D、既没有登记性误差，也没有代表性误差

4. 要使抽样误差减少一半 (在其它条件不变的情况下)，则抽样单位数必须 ()。

- A、增加 2 倍 B、增加到 2 倍 C、增加 4 倍 D、增加到 4 倍

5. 抽样平均误差反映了样本指标与总体指标之间的 ()。

- A、实际误差 B、实际误差的绝对值
- C、平均误差程度 D、可能误差范围

6. 反映样本指标与总体指标之间抽样误差可能范围的指标是 ()。

- A、样本平均误差 B、抽样极限误差
- C、可靠程度 D、概率度

7. 在实际工作中，不重复抽样的抽样平均误差的计算，采用重复抽样的公式的场合是

()

- A、抽样单位数占总体单位数的比重很小时
- B、抽样单位数占总体单位数的比重很大时
- C、抽样单位数目很少时 D、抽样单位数目很多时

8. 在其它条件不变的情况下，抽样单位数目和抽样误差的关系是 ()。

- A、抽样单位数目越大，抽样误差越大
- B、抽样单位数目越大，抽样误差越小

C、抽样单位数目的变化与抽样误差的数值无关

D、抽样误差变化程度是抽样单位数变动程度的 1/2

9.事先将全及总体各单位按某一标志排列，然后依固定顺序和间隔来抽选调查单位的抽样组织形式，被称为（）。

A、分层抽样 B、简单随机抽样 C、整群抽样 D、等距抽样

10.在总体内部情况复杂，而且各单位之间差异程度较大，单位数目又多的情况下，宜采用（）。

A、简单随机抽样 B、类型抽样 C、等距抽样 D、整群抽样

11.用考虑顺序的重置抽样方法，从 4 个单位中抽选 2 个单位组成一个样本，则样本可能数目为（）。

A、 $4^2 = 16$ B、 $\frac{5!}{2!3!} = 10$ C、 $\frac{4!}{2!} = 12$ D、 $\frac{4!}{2!2!} = 6$

12.无偏性是用抽样指标估计总体指标应满足的要求之一，无偏性是指（）。

A、样本平均数等于总体平均数

B、样本成数等于总体成数

C、抽样指标等于总体指标

D、抽样指标的平均数等于总体指标

13.在同样条件下，不重置抽样的抽样平均误差与重置抽样的抽样平均误差相比（）。

A、前者小于后者 B、前者大于后者

C、两者相等 D、无法判断

14.当可靠程度大于 0.6828 时，抽样极限误差（）。

A、大于抽样平均误差 B、小于抽样平均误差

C、等于抽样平均误差

D、与抽样平均误差的大小依样本容量而定

15.在其它条件不变的情况下，提高抽样估计的可靠程度，其精确度将（）。

A、保持不变 B、随之扩大 C、随之缩小 D、无法确定

16.如果总体成数方差未知，计算必要抽样数目时，可用总体方差的最大值，最大值为（）。

A、0.24 B、0.25 C、0.50 D、1

(五) 多项选择题

1. 抽样调查适用于下列哪些场合 ()。

- A、不宜进行全面调查而又要了解全面情况
- B、工业产品质量检查 C、调查项目多、时效性强
- D、只需了解一部分单位的情况 E、适用于任何调查

2. 下列哪些调查中存在系统误差 ()。

- A、抽查某车间产品质量时将两名徒工的产品排除在外
- B、某县用等距抽样调查农户收入时决定：如果中选户户主不在家，则以该户下一户顶替，已知不在家的农户多为外出经商者

- C、为了了解学生健康情况而做的普查
- D、将学生按来自城乡分组后采取按比例抽样调查学生开支情况
- E、简单随机抽样

3. 影响抽样误差大小的因素有 ()。

- A、抽样单位数的多少 B、总体被研究标志的变异程度
- C、抽样组织方式 D、登记误差大小
- E、抽到的样本如何

4. 合理的或优良的抽样估计，要求用抽样指标估计总体指标时，应该满足如下标准，即 ()。

- A、无偏性 B、代表性 C、一致性 D、时间性
- E、有效性

5. 抽样推断中的抽样误差 ()。

- A、是不可避免要产生的 B、可以事先计算
- C、可通过不同的抽样方式消除
- D、误差范围可事先控制 E、只能事后进行计算

6. 抽样平均误差是 ()。

- A、所有可能样本平均数的标准差
- B、所有可能样本平均数的平均差
- C、抽样推断中作为计算误差范围的衡量尺度
- D、反映抽样平均数与总体平均数的平均误差程度

E、是一种系统性误差

7.区间估计的基本要素有（）。

A、平均误差 B、概率度 C、修正系数 D、点估计值

E、误差范围

8.抽样调查的数理依据主要有（）。

A、回归分析 B、大数定律 C、中心极限定理

D、误差分布理论 E、相关分析

9、与典型调查相比较，抽样调查的特点有（）。

A、理论基础是概率论大数法则

B、按随机原则抽选调查单位

C、抽样误差可以事先计算和控制

D、调查单位的选取方法不同

E、排除主观因素的影响

10.抽样估计的特点是（）。

A、在逻辑上运用演绎推理 B、在逻辑上运用归纳推理

C、在方法上运用不确定的概率估计法

D、在方法上运用确定的数学分析法

E、结论存在一定的抽样误差

(六) 简答题

1.什么是随机性原则？在抽样调查中为什么要坚持随机性原则？

2.什么是抽样估计？它有什么特点？

3.什么是抽样误差？影响抽样误差的因素有哪些？

4.什么叫点估计？什么叫区间估计？

5.什么是极限误差，它与概率度，抽样平均误差有什么关系？

(七) 论述题

重复抽样与不重复抽样有什么不同？为什么不重复抽样的误差总是小于重复抽样的误差？

(八) 计算题

1.某商店有 20 名职工，现从中抽取 5 名职工进行调查。采用简单随机抽样，试计算各

种抽样方法的样本可能数目。

2.对某产品进行重量测试,被抽中的10袋产品其重量如下(单位:克):
48,47,49,50,51,48,46,45,46,49 试计算产品重量的抽样平均误差。

3.某灯泡厂生产一批灯泡共8000只,随机抽选400只进行耐用时间的试验。测试结果平均寿命5000小时,总体标准差为300小时。求抽样误差。

4.某市在家庭平均人数调查中,采用不重置简单随机抽样得到样本的每户平均人数为3.2人,标准差为1.378人。

① 如果总户数为1000户,调查户数为80户,试计算抽样的平均误差;

② 在①条件下,如果极限误差为0.2956,试估计户平均人数;

③ 在①的条件下,如果可靠程度为0.95,试计算户平均人数;

④ 如果要求可靠程度为95.45%,极限误差为0.35,试计算重置和不重置抽样各需抽取多少户进行调查(总户数仍为1000户)。

5.某机械厂生产一批零件共6000件,随机抽查300件,发现其中有9件不合格,求合格率的抽样误差。

6.某服装厂对当月生产的20000件衬衫进行质量检查,结果在抽查的200件衬衫中有10件是不合格品,要求:

(1) 以95.45%概率推算该产品的合格范围;

(2) 该月生产的产品是否超过规定的8%的不合格率(概率不变)。

7.某地农村种植小麦150亩,在麦收前随机不重置抽取了100个平方公尺的小麦样本,测得每平方公尺小麦产量为0.5公斤,标准差为0.05公斤。试计算极限抽样误差,并以95.45%的概率保证,推断该地区小麦平均亩产量和总产量。

8.某茶叶公司销售一批茗茶,规定每包规格重量不低于150克,现抽取1%检验,结果是:

按每包重量分组(克)	包数
148~149	10
149~150	20
150~151	50
151~152	20
合计	100

试以99.73%的概率,评估这批茶叶平均每包重量的范围是否符合规定重量要求。

9.某厂 4500 名职工中，随机抽取 20%调查每月看电影次数，所得资料如下：

看电影次数	0 ~	2 ~	4 ~	6 ~	8 ~
	2	4	6	8	10
职工人数 (对总数的百分比)	8	22	40	25	5

试以 95.45%的可靠性：（1）估计平均每月看电影的次数；（2）确定每月看电影在 4 次以上的比重，其误差不超过 3%。

10.年终在某乡镇储蓄所中按定期储蓄存款账号进行抽样调查，得到如下资料：

存款额	抽查户数
1000 元以	58
下	150
1000 ~	200
2000	62
2000 ~	14
3000	
3000 ~	
4000	
4000 元以	
上	
	484

（1）试估计该储蓄所所有储户平均存款额的区间（概率为 95.45%）；

（2）估计定期存款额在 2000 元以上的比重。

11.工商部门对某超市经销的小包装休闲食品进行重量合格抽查，规定每包重量不低于 30 克，在 1000 包食品中抽 1%进行检验，结果如下表：

按重量分组 (克)	包数 (包)
26~27	1
27~28	3
28~29	3
29~30	2
30~31	1
合计	10

--	--

试以 95.45% 概率推算：

- (1) 这批食品平均每包重量是否符合规定要求；
- (2) 若每包食品重量低于 30 克为不合格，求合格率的范围。

12. 对某居委会 30 户家庭的月收支情况进行抽样调查，发现平均每户每月用于书报费支出为 45 元，抽样平均误差为 2 元，试问应以多少概率才能保证每户每月书报费支出在 41.08 元至 48.92 元之间。