项目 3

选择抽样方式

项目导言

在市场调查中，为了取得某一市场的总体情况，运用全面调查方法可以取得全面、完整的统计资料，进而了解市场总体特征。但在许多情况下，比如当市场总体非常大、总体单位数非常多或者市场总体的综合特征要经过破坏性测试才能取得的情况下，对总体单位进行全面调查非常困难，也根本不可能，这时只能对部分单位进行抽样调查，进而推断总体的综合特征。在市场调查中，抽样调查作为一种非全面调查方式已经成为一种非常重要且应用广泛的调查方式。

学习目标

- **能力目标**

1. 熟练运用抽样调查的各种方式抽取样本，并会计算样本数目和抽样误差。
2. 能够根据调研课题的需要设计合理的抽样方案。

- **知识目标**

1. 了解抽样调查的概念和特点。
2. 掌握抽样调查的程序以及抽样调查的各种方式和方法。

- **素质目标**

1. 培养学生数理分析思维。
2. 培养学生自主学习、自我管理能力和团队合作精神。

案例导入

国务院决定于 2015 年开展全国 1%人口抽样调查

据中国政府网消息，国务院日前下发《国务院办公厅关于开展 2015 年全国 1%人口抽样调查的通知》(以下简称《通知》)，决定于 2015 年开展全国 1%人口抽样调查，了解 2010 年以来中国人口在数量、素质、结构、分布以及居住等方面的变化情况，为制定国民经济和社会发展规划提供科学、准确的统计信息支持。

《通知》显示，在中国境内抽取约 6 万个调查小区，调查对象为小区内的全部人口(不包括港澳台居民和外国人)，共约 1400 万人。调查内容为人口和住户的基本情况，主要包括姓名、性别、年龄、民族、受教育程度、行业、职业、迁移流动、社会保障、婚姻、生育、死亡、住房情况等。调查时点为 2015 年 11 月 1 日零时。

《通知》要求，要按照“统一领导、分工协作、分级负责、共同参与”的原则，做好调查的组织和实施工作。为加强领导和协调，由统计局会同有关部门成立2015年全国1%人口抽样调查工作协调小组（以下简称协调小组）。协调小组办公室设在统计局，负责调查的组织实施和日常工作，督促落实协调小组议定事项。发展改革部门负责做好调查方案与国民经济和社会发展总体规划及有关专项规划编制实施的衔接；公安部门负责提供各级户籍人口、流动人口等资料并协助做好现场登记；宣传部门负责做好新闻宣传，以及新闻媒体的组织协调。其他部门按照职能分工，认真做好相关工作。县级以上地方各级人民政府要切实加强组织领导，建立相应机构，确保调查任务顺利完成。

《通知》显示，2015年全国1%人口抽样调查所需经费，按照分级负担原则，由中央和地方各级人民政府共同负担，并列入相应年度的财政预算，按时拨付、确保到位。

《通知》要求，要坚持依法调查。要严格执行《中华人民共和国统计法》和《全国人口普查条例》的有关规定。调查取得的数据，严格限定用于调查目的，不得作为任何部门和单位对各级行政管理工作实施考核、奖惩的依据，不得作为对调查对象实施处罚的依据；各级调查机构及其工作人员，必须严格履行保密义务。做好宣传引导。要通过报刊、广播、电视和网络等方式，广泛深入宣传调查的重要意义和工作要求，引导广大调查对象依法配合调查，如实申报调查项目，为调查工作顺利实施创造良好舆论环境。

（资料来源：中国新闻网. http://www.chinanews.com/gn/2014/07-07/6358178.shtml）

案例思考

1. 抽样调查的应用范围和现实意义是什么？
2. 如何实施抽样调查？

通过以上案例，我们可以得到这样的结论：抽样调查是一种非全面调查，它是从全部调查研究对象中，抽选一部分单位进行调查，并据以对全部调查研究对象做出估计和推断的一种调查方法。抽样调查经济性好、实效性强、适用面广、准确性较高，被认为是非全面调查方法中用来推算和代表总体的最普遍、最有科学根据的调查方法。为了保证样本的代表性，降低抽样误差，抽样必须遵循正确的程序和规则。抽样设计是市场调研中非常重要的一个环节，它是影响调研结论是否有效的重要因素之一。因此，要全面认知抽样调查，掌握抽样调查的工作流程、方式和方法，并学会设计合理的抽样方案。

任务1 认知抽样调查

任务引入

各公司结合所选调研课题，进行抽样设计，并完成具体的抽样设计方案。

- 问题1：什么是抽样调查？
- 问题2：抽样的方式、方法有哪些？

知识铺垫

3.1.1 抽样调查的概念、特点和分类

1. 抽样调查的概念

抽样调查实际上是一种专门组织的非全面调查。它是按照一定方式，从调查总体中抽取部分样本进行调查，用所得的结果说明总体情况的调查方法。抽样调查是现代市场调查中的重要组织形式，是目前国际上公认和普遍采用的科学的调查手段。抽样调查的理论原理是概率论，概率论中诸如中心极限原理等一系列理论，为抽样调查提供了科学的依据。

2. 抽样调查的特点

① 从经济上说，抽样调查节约人力、物力和财力；

② 抽样调查更节省时间，具有较强的时效性；

③ 抽样调查具有较强的准确性；

④ 通过抽样调查，资料收集的深度和广度都会大大提高。

尽管抽样调查具有上述优点，但它也存在着某些局限性。它通常只能提供总体的一般资料，而缺少详细的分类资料，在一定程度上难以满足对市场经济活动分析的需要。此外，当抽样数目不足时，将会影响调查结果的准确性。

3. 抽样调查的分类

抽样调查总体上分为随机抽样和非随机抽样两大类。

随机抽样是按照随机原则抽取样本，即在总体中抽取单位时，完全排除了人的主观因素的影响，使每一个单位都有同样被抽到的可能性。遵守随机原则，一方面，可使抽取出来的部分单位的分布情况（如不同年龄、文化程度人员的比例等）有较大的可能性接近总体的分布情况，从而使根据样本所做出的结论对总体研究具有充分的代表性；另一方面，遵循随机原则，可有助于调查人员准确地计算抽样误差，并有效地加以控制，从而提高调查的精度。

非随机抽样不遵循随机原则，它是从便利性出发或根据主观的选择来抽取样本。非随机抽样无法估计和控制抽样误差，无法用样本的定量资料采用统计方法来推断总体。但非随机抽样简单易行，尤其适用于做探测性研究。

4. 抽样调查的适用范围

① 对一些不可能或不必要进行全面调查的社会经济现象，最宜用抽样方式解决，例如对有破坏性或损耗性质的商品进行质量检验、对一些具有无限总体的调查（如对森林木材积蓄量的调查）等。

② 在经费、人力、物力和时间有限的情况下，采用抽样调查的方法可节省费用，争取时效，用较少的人力、物力和时间达到满意的调查效果。

③ 运用抽样调查对全面调查进行验证，全面调查涉及面广、工作量大、花费时间和经费多，组织起来比较困难。但调查质量如何需要检查验证，这时，显然不能用全面调查方式进行。例如，工业普查前后需要几年的时间才能完成，为了节省时间和费用，常用抽样调查进

行检查和验证。

④ 对某种总体的假设进行检验，判断这种假设的真伪，以决定行为的取舍时，也经常用抽样调查来测定。

3.1.2 抽样调查中的常用概念

1. 总体与样本

总体是所要调查研究的现象的全体，它是由具有同质性和差异性的许多个别事物的集合体。总体单位数通常用 N 表示。

样本是按随机原则从总体中抽出来的一部分单位的综合体，样本中包含的单位个数称为样本量，用 n 表示。n/N 称为抽样比。

例如，某学校要调查学生的后勤服务满意度，全校共有学生 10000 人，从中抽取 500 人进行调查，那么总体 N 就是该校的全部学生数，即 $N=10000$；样本 n 就是所要抽取的那部分学生，即 $n=500$；抽样比为 $n/N=500/10000=5\%$。

2. 参数与统计量

参数是总体的数量特征，即总体指标。参数在抽样时往往是未知的，是需要进行推断的。参数通常有总体均值($\overline{X}$)、总体标准差(σ)、总体比率(P)等。

统计量是样本的数量特征，即样本指标。统计量随样本不同而不同，因而是一个随机变量。统计量通常有样本均值($\bar{x}$)、样本标准差(S)、样本比率(p)等。

3. 抽样框与抽样单位

抽样框是一个包括全部总体单位的框架，用来代表总体，以便从中抽取样本的一个框架。抽样框可以是一览表(如名单或名录)、一本名册、一幅地图、一段时间等。

抽样单位是指样本抽取过程中的单位形式，也即从抽样框中直接抽取的单位称为抽样单位，它可能是总体中的基本单位，也可能是总体中的基本单位的集合。

例如，欲调查某市大学的教学用品需求，则全市大学的集合为总体，抽样框是全市的大学名单。总体单位是每一个大学，抽样单位可以是总体中的每一个大学，也可以是大学分类中的每一个大学。

4. 样本量与样本单位

样本量是指样本的大小，即一个样本中包含的样本单位的多少。样本量的大小，取决于抽样调查的精度要求、总体各单位的标志变异程度、抽样估计的可信程度、抽样方式方法等因素的制约。

样本单位是构成样本的基本单位，与总体单位的形式是一致的，样本单位可以直接从总体中抽取，也可从抽样单位中产生。

5. 总体分布、样本分布与抽样分布

总体分布：总体各单位标志值的分布状况，又称总体结构。

样本分布：样本中各样本单位标志值的分布状况，又称样本结构。当样本量足够大时，样本分布趋于总体分布。

抽样分布：从总体中抽取的所有可能的样本的统计量构成的分布。根据中心极限定量，当样本量足够大时，样本均值等统计量的分布趋近于正态分布，因而可用正态分布来作区间估计。

6. 重复抽样与不重复抽样

从 N 个总体单位中抽取 n 个组成样本，有两种抽取方法。

① 重复抽样，即每抽出一个单位进行登记后，放回去，混合均匀后，再抽下一个，直到抽满 n 个为止。重复抽样有可能出现极大值或极小值组成的极端样本。

② 不重复抽样，即每次抽出一个单位进行登记后，不再放回参加下一次抽取，依次下去，直到抽满 n 个为止。不重复抽样可以避免极端样本出现，抽样误差比重复抽样小。

7. 抽样误差与抽样标准误差

抽样误差是指在遵守随机原则条件下，样本指标与总体指标之间的差异，它是一种偶然性的代表性误差，不包括系统性误差和非抽样误差。抽样误差的大小通常受样本量大小、总体标准差、抽样方法、抽样方式四个因素的影响。

抽样误差的大小常用抽样标准误差来反映，而抽样标准误差是指所有可能的样本均值（或样本比率）与总体均值（或总体比率）的标准差，抽样标准误差的平方称为抽样方差。依定义有：

$$\sigma_{\bar{x}}=\sqrt{\frac{\sum(\bar{x}-\overline{X})^2}{N}}$$

$$\sigma_{p}=\sqrt{\frac{\sum(p-P)^2}{N}}$$

式中，$\sigma_{\bar{x}}$ 代表样本平均数的抽样标准误差，σ_{p} 代表样本比率的抽样标准误差；N 代表样本个数。上述公式可用来解释抽样误差的实质，但不能实际应用，因为所有可能的样本个数太多，总体均值或总体比率是未知的，是需要推断的，同时，实际抽样时，往往只能抽取一个样本进行调查。因此，抽样标准误差的计算需要寻求别的测定方法，将在以下各种抽样方式中介绍。

8. 点估计与区间估计

点估计也叫定值估计，当样本容量足够大时，可直接用样本均值代替总体均值，用样本比率代替总体比率，可据此计算有关总量指标，这就是点估计。

区间估计是用一个取值区间及其出现的概率来估计总体参数。具体来说，区间估计是用样本统计量和抽样标准误差来构造总体参数的取值范围，并用一定的概率来保证总体参数落在估计的区间内。其概率称为置信概率，概率的保证程度称为可靠性或置信度（t），估计区间称为置信区间。如：

总体均值

$$\overline{X}=\bar{x}\pm t\sigma_{\bar{x}}$$

总体比率

$$P=p\pm t\sigma_{\bar{p}}$$

式中，$t\sigma_{\bar{x}}$ 和 $t\sigma_{\bar{p}}$ 又称为允许误差或极限误差，记作 Δ，$\Delta/\overline{X}$，Δ/P 称为估计的相对精度。

中心极限定理已证明，概率度 t 和概率 p 呈函数关系，即 $p=F(t)$，t 每取一个值，都有唯一

确定的 p 值与之相对应。在实际工作中，为了使用方便，将不同的 t 值与其相应的概率 p 预先算好，编成概率表，供调查时使用。几个常用的概率度和概率之间的关系如表 3-1 所示。

表 3-1 概率度和概率函数关系

t	$F(t)$	t	$F(t)$
1.00	0.6827	2.50	0.9876
1.50	0.8664	3.00	0.9973
1.96	0.9500	4.00	0.9994
2.00	0.9545	5.00	0.999999

9. 抽样方式与抽样方法

① 抽样方式是指抽样调查的组织方式，通常有简单随机抽样、分层抽样、系统抽样、整群抽样、目录抽样、多阶段抽样等。这些抽样调查的组织方式、抽样误差的计算和区间估计，下面将分别介绍。

② 抽样方法是指在抽样调查的组织方式既定的前提下，从总体的全部单位（个体）中抽取 n 个单位组成样本的方法。通常有重复抽样与不重复抽样两种抽取方法，而重复抽样与不重复抽样的具体实施，又有不同的具体做法。

3.1.3 抽样调查方式

1. 随机抽样

1）简单随机抽样

（1）简单随机抽样的概念

简单随机抽样也称纯随机抽样，是指在总体单位均匀混合的情况下，随机逐个抽出样本的抽样方式，它是概率抽样的最基本类型。简单随机抽样只适用于总体单位数不多，总体单位标志变异度较小的情形。通常采用直接抽取法、抽签法、随机数表法等抽取样本。

① 直接抽取法是从调查总体中直接随机抽取样本进行调查。这种方法适合对集中在某个较小空间的总体进行抽查。例如，对存放在仓库中的所有同类产品随机抽查出若干箱产品为样本进行质量检验。但这种方法有难以完全遵循随机的缺点，因为在抽选的过程中往往受到主观判断的影响，所以采取这种方法时避免主观判断的影响是关键。在正式调查中，很少采用直接抽选法。

② 抽签法是先将研究总体中的每一个单位统一编号，使每一个单位都有一个号，然后将每一个号做成一个卡号并且混合均匀，最后从中随机抽取卡片，直到抽到额定的数目为止。抽签法有重复抽样和不重复抽样两种方式，这种方法在日常生活中用得比较多。

【例 3-1】 要从 500 名学生中抽取 50 人进行调查，采用抽签法如何抽取样本？

首先把这 500 名学生的姓名分别写在小纸条上，再把 500 张小纸条放在一个纸箱中摇匀，然后任意抽取一张，则该学生就是样本的第一个单位。依次取出 50 张，就构成此次抽样的样本，这是不重复抽样。如果每一次都把抽取出的纸条放回去，再任意抽出，出现重复的则再放回去抽取一次，直至抽到 50 个不同的学生姓名，这就是重复抽样。

③ 随机数表又称为乱数表，它是将 0～9 的 10 个自然数，按编码位数的要求（如两位一

组、三位一组、五位甚至十位一组），利用特制的摇码器（或电子计算机），自动地逐个摇出（或电子计算机生成）一定数目的号码编成表，以备查用。这个表内任何号码都有同等出现的可能性。利用这个表抽取样本时，可以大大简化抽样的烦琐程序。其缺点是不适用于总体中个体数目较多的情况。随机数表法应用的具体步骤是：将调查总体单位一一编号；在随机号码表上任意规定抽样的起点和抽样的顺序；依次从随机号码表上抽取样本单位号码。凡是抽到编号范围内的号码，就是样本单位的号码，一直到抽满为止。采用随机号码表法抽取样本，完全排除主观挑选样本的可能性，使抽样调查有较强的科学性。

【例 3-2】 某企业调查消费者对某产品的需求量，要从 90 户居民家庭中抽选 10 户居民，采用随机数表法如何抽选样本？

具体步骤如下。

第一步：将 90 户居民家庭编号，每一户家庭一个编号，即 01～90。（每户居民编号为 2 位数）

第二步：在如表 3-2 所示随机数中，随机确定抽样的起点和抽样的顺序。假定从第一行第 6 列开始抽，即从号码“36”作为起始号码，抽样顺序从左往右抽。

表 3-2 随机数（片段）

03	47	43	73	86	36	96	47	36	61	46	99	69	81	62
97	74	24	67	62	42	81	14	57	20	42	53	32	37	32
16	76	02	27	66	56	50	26	71	07	32	90	79	78	53
12	56	85	99	26	96	96	68	27	31	05	03	72	93	15
55	59	56	35	64	38	54	82	46	22	31	62	43	09	90
16	22	77	94	39	49	54	43	54	82	17	37	93	23	78
84	42	17	53	31	57	24	55	06	88	77	04	74	47	67
63	01	63	78	59	16	95	55	67	19	98	10	50	71	75
33	21	12	34	29	78	64	56	07	82	52	42	07	44	28
57	60	86	32	44	09	47	27	96	54	49	17	46	09	62

第三步：从起始号码开始，从左到右依次抽取 10 个不重复的位于 01～90 的号码，由此产生的 10 个样本单位号码为：36、47、61、46、69、81、62、74、24、67。编号为这些号码的居民家庭就是抽样调查的对象。

（2）简单随机抽样标准误差

① 样本平均数的抽样标准误差为：

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} \quad \text{（重复抽样）}$$

或

$$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)} \quad \text{（不重复抽样）}$$

② 样本比率的抽样标准误差为：

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} \quad \text{（重复抽样）}$$

或

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \quad \text{（不重复抽样）}$$

【例 3-3】 某商场从某天的顾客中，不重复随机抽取 100 个顾客调查购买商品情况，其中有 5 个顾客未购买商品(未购率 5%)；顾客购买商品的样本平均数为 498 元，样本标准差为 144 元，要求用 95%的概率($t=1.96$)估计顾客平均购买额和未购率的置信区间。

解：此题不知总体方差，因样本为大样本，可用样本方差代替。

$$\sigma_{\bar{x}} = \sqrt{\frac{144^2}{100}} = 14.4$$

$$\sigma_p = \sqrt{\frac{0.05(1-0.05)}{100}} = 0.022$$

顾客平均购买额的置信区间为：

$$498 \pm 1.96 \times 14.4 \quad \text{即}[469.78, 526.22]\text{元}$$

顾客未购率的置信区间为：

$$5\% \pm 1.96 \times 2.2\% \quad \text{即}[0.69\%, 9.31\%]$$

(3) 简单随机抽样样本容量的确定

一般来说，样本容量确定应考虑总体方差 σ^2、抽样估计精度要求(允许误差 Δ 的约束)和把握程度(置信概率)的大小、抽样方式方法、抽样调查费用约束等因素。在不考虑抽样调查费用约束的条件下，样本容量的计算公式为：

① 总体均值估计所需的样本容量

$$n = \frac{t^2\sigma^2}{\Delta^2} \quad \text{(重复抽样)}$$

或

$$n = \frac{t^2\sigma^2 N}{N\Delta^2 + t^2\sigma^2} \quad \text{(不重复抽样)}$$

② 总体比率估计所需的样本容量

$$n = \frac{t^2 p(1-p)}{\Delta^2} \quad \text{(重复抽样)}$$

或

$$n = \frac{Nt^2 P(1-P)}{N\Delta^2 + t^2 P(1-P)} \quad \text{(不重复抽样)}$$

(4) 简单随机抽样的优缺点和适用范围

简单随机抽样只适用于总体单位数量有限的情况，否则编号工作繁重；对于复杂的总体，样本的代表性难以保证；不能有效地利用总体的已知信息等。在市场调研范围有限，或调查对象情况不明、难以分类，或总体单位之间特性差异程度小的情况下采用此法效果较好。

2) 分层抽样

(1) 分层抽样的概念

分层抽样又称为类型抽样，是先将总体按有关的研究标志分组，然后再从每组中按随机原则抽取样本。在每个组中抽取的调查单位的数目，可按相同的比例(n/N)抽取，也可按不同的比例抽取。为了简便起见，通常都是按相同比例抽取，称作等比例分层抽样。

在分层抽样时，抽样误差只和层内方差有关，而与层间方差无关。因此，只要能够扩大层间方差而缩小层内方差，就可以提高抽样效率。

(2) 分层抽样的基本步骤

① 分层。将总体按照一定的标准进行分层，选择分层标准时要注意：分层后，同一层内

部的单位尽可能是同质的，不同层之间的单位尽可能是异质的。

② 确定各层所要抽取的样本量。

具体做法有以下两种。

第一种：等比例分层抽样。即在确定各层所要抽取的样本数量时，按各层占总体的比例分配各层的样本数量。

【例 3-4】 某公司要估计某地家用电器的潜在用户。这种商品的消费同居民收入水平相关，因而以家庭年收入为分层基础。假定某地居民为 100000 户，已确定样本数为 1000 户，家庭年收入在 10000 元以下的家庭户数为 18000 户，收入在 10000～30000 元的家庭户数为 35000 户，收入在 30000～60000 元的家庭户数为 30000 户，收入在 60000 元以上的家庭户数为 17000 户，请采用等比例分层抽样法确定各层应抽取的样本数。

$$10000\text{ 元以下抽取样本数}=1000\times\frac{18000}{100000}=180(\text{户})$$

$$10000\sim30000\text{ 元抽取样本数}=1000\times\frac{35000}{100000}=350(\text{户})$$

$$30000\sim60000\text{ 元抽取样本数}=1000\times\frac{30000}{100000}=300(\text{户})$$

$$60000\text{ 元以上抽取样本数}=1000\times\frac{17000}{100000}=170(\text{户})$$

第二种：不等比例分层抽样，又称分层最佳抽样。这种抽样法不按各层中样本单位数占总体单位数的比例分配各层样本数，而是根据各层的标准差的大小来调整各层样本数目。该方法既考虑了各层在总体中所占比重的大小，又考虑了各层标准差的差异程度，有利于降低各层的差异，以提高样本的可信程度，故也可将不等比例分层抽样称为分层信任程度抽样。

【例 3-5】 某公司要调研某地家用电器产品的潜在用户，这种产品的消费同居民收入水平有关，因此以家庭收入为分层基础。假定该地居民户即总体单位数为 20000 户，已确定调研样本数为 200 户。家庭收入分高、中、低三层，其中高档收入家庭为 2000 户，占总体单位数的 10%；中等收入家庭为 6000 户，占总体单位数的 30%；低等收入家庭为 12000 户，占总体单位数的 60%。现又假定各层样本标准差为：高档收入家庭是 300 元，中等收入家庭是 200 元，低等收入家庭是 50 元。现要求根据分层最佳抽样法，确定各收入层家庭应抽取的户数。

为了便于观察，列表 3-3 计算如下。

表 3-3 调研单位数与样本标准差乘积计算

家庭收入分层	各层调研单位数	各层的样本标准差	乘积	样本单位数
高	2000	300	600000	$200\times\frac{600000}{2400000}=50$
中	6000	200	1200000	$200\times\frac{1200000}{2400000}=100$
低	12000	50	600000	$200\times\frac{600000}{2400000}=50$
合计	20000		2400000	200

如果按照等比例分层抽样，那么，高档收入家庭的分层样本数为 20 户（200×10%）；中等收入家庭的分层样本数为 60 户（200×30%）；低等收入家庭的分层样本数为 120 户（200×60%）。将前后两种方法抽取的各层样本数做个对比，不难看出，相比于等比例分层抽样法，根据分层最佳抽样法抽取样本，则高档收入家庭的分层样本数增加了 30 户，中等收入家庭的分层样本数增加了 40 户，低等收入家庭的分层样本数则减少了 70 户。由于购买家用电器同家庭收入水平是呈正比例变动的，所以，增加高、中档层的样本数，相应减少低档层的样本数，将有利于提高抽样的准确性。

③ 在各层内部进行抽样，即按照随机原则，用简单随机抽样或等距抽样的方法，从各层中抽取所要的样本数目，各层的样本之和构成了样本总体。

（3）分层抽样的抽样标准误差

设 n_i、$\overline{x_i}$、σ_i^2 分别为样本各组的单位数、平均数、方差；N_i 为总体各组的单位数，在等比例分层抽样条件下，则有下列计算公式。

总体平均数点估计

$$\overline{X}=\frac{\sum \overline{x_i}N_i}{\sum N_i}=\frac{\sum \overline{x_i}n_i}{\sum n_i}$$

层内方差平均数

$$\overline{\sigma^2}=\frac{\sum \sigma_i^2 N_i}{\sum N_i}=\frac{\sum \sigma_i^2 n_i}{\sum n_i}$$

总体平均数的抽样标准误差

$$\sigma_{\overline{x}}=\sqrt{\frac{\overline{\sigma^2}}{n}}\quad\text{（重复抽样）}$$

或

$$\sigma_{\overline{x}}=\sqrt{\frac{\overline{\sigma^2}}{n}\left(1-\frac{n}{N}\right)}\quad\text{（不重复抽样）}$$

总体比率估计的抽样误差的计算只需用 $p_i(1-p_i)$ 代替上述层内方差平均数公式中的 σ_i^2 即可；而总体比率估计的公式为：

$$P=\frac{\sum p_i N_i}{\sum N_i}=\frac{\sum p_i n_i}{\sum n_i}$$

【例 3-6】 某县某年共有乡镇 18 个，农民家庭 88 万户，按各乡镇收入高低可分为高收入乡镇、中收入乡镇、低收入乡镇三类，各类乡镇的农户数如表 3-4 所示，现从这三个类别中按等比例抽样，共抽取 500 户组成样本，要求在 90% 的置信概率（$t=1.64$）下对全县户均年收入进行区间估计。

表 3-4 某市居民收入分层抽样数据

类 型	家庭 N_i（万户）	样本容量 n_i（户）	户均年收入 $\bar{x}_i$（百元）	标准差 σ_i
高收入	38.72	220	700	200
中收入	31.68	180	400	120
低收入	17.60	100	300	180
合 计	88.00	500	—	—