



第二章 土地统计调查

第一节 概述

第二节 土地详查

第三节 土地变更统计调查

第四节 土地质量统计调查

第五节 土地抽样调查

第五节 土地抽样调查

- 一、土地抽样调查的基本原理
- 二、土地抽样调查的组织形式
- 三、抽样误差
- 四、总体指标估计
- 五、样本容量的确定

一、土地抽样调查的基本原理

- **抽样调查**：按**随机**原则从全部研究单位中抽取一部分单位进行观察，根据样本资料计算**样本的特征值**，据以对**总体的特征值**做出具有一定可靠性的估计和判断，以反映总体的数量特征和数量表现的一种统计方法。
- **随机原则**：在抽取样本时，排除人们主观意图的作用，使得总体中的各单位均以相等的机会被抽中。又称为**等可能性原则**。
- **例**：（1）小麦测产。（2）产品质量合格率检查。

(一) 抽样调查的基本特点

- 1、调查单位的确定是按随机原则从全部总体单位中抽取的。
- 2、用部分单位的指标数值去推断和估计总体指标数值。
- 3、抽样调查中的抽样误差是不可避免的，但在事先是可以计算并加以控制的。

(二) 抽样调查的应用范围

- 1、**无法**全面调查的现象，为测算全面资料，采用抽样调查。如土地质量调查。
- 2、**没有必要或很难**全面调查的现象，采用抽样调查。如全国城乡人民家庭生活状况调查。抽样调查节约时间、人力、物力和财力，提高时效性。
- 3、抽样调查结果可对全面调查结果进行**检查和修正**。全面调查之后抽样复查，计算差错率，据以检查和修正全面调查结果，提高全面调查质量。
- 4、要求及费用高的调查，如**土地现象**，可采用抽样调查。如土壤养分调查。

(三) 抽样调查的几个基本概念

- 1. 总体和样本
- 2. 总体指标和样本指标
- 3. 重复抽样和不重复抽样

1. 总体和样本

- **总体/全及总体或母体**：指所要认识的研究对象的全体，它是由所研究范围内具有某种**共同性质**的全体单位所组成的集合体。用N代表全及总体的单位数。
- **样本/抽样总体**：按**随机**原则从总体中**抽取**的一部分单位组成的小总体。它也是由许多性质相同的单位组成的。用n代表样本的单位数， n 也称为**样本容量**。组成样本的每个单位称为**样本单位**。
- 按照样本容量的大小样本可分为**大样本和小样本**。 $n \geq 30$ 为大样本， $n < 30$ 为小样本。抽样调查时，多数采用大样本。
- **注意**：作为抽样推断对象的**总体是唯一**确定的，但作为观察对象的**样本不是唯一的**。从一个总体中可以抽取很多个样本，每次抽到哪个样本是不确定的。

2. 总体指标和样本指标

□ 总体指标

□ 主要有：

- ① 平均指标
- ② 标准差（或方差）指标
- ③ 成数指标

□ 样本指标

□ 对应指标：

- ① 平均指标
- ② 标准差（或方差）指标
- ③ 成数指标

3. 重复抽样和不重复抽样

- (1) **重复抽样**：是从总体中抽取样本时，随机抽取一个样本单位，记录该单位有关标志表现以后，把它**放回**到总体中去，再从总体中随机抽取第二个单位，记录它有关标志表现以后，也把它放回全及总体中去，照此下去直到抽选 n 个样本单位。
- 可见，重复抽样时全及**总体单位数**在抽选过程中始终**没有减少**，而且**各单位有被重复抽中的可能**。

3. 重复抽样和不重复抽样

- (2) **不重复抽样**：从总体中抽取第一个样本单位，记录该单位有关标志表现后，该样本单位**不再放回**总体中参加下一次抽选。然后，从总体 $N-1$ 个单位中随机抽选第二个样本单位，记录该单位标志表现后，该单位也不再放回总体，再从总体 $N-2$ 单位中抽选第三个样本单位，照此下去直到抽选出 n 个样本单位。
- 可见，不重复抽样时，**总体单位数**在抽选过程中是**逐渐减少**的，而且**各单位没有重复被抽中可能**。
- 两种抽样方法产生三个差别：①抽取的样本可能数目不同；②抽样误差计算公式不同；③抽样误差的大小不同

二、土地抽样调查的组织形式

- 1、简单随机抽样/纯随机抽样
- 按随机原则直接从总体 N 个单位中抽取 n 个单位作样本，使每个总体单位都有同等的机会被抽中。是抽样调查中最基本的、最单纯的方式，适于均匀总体。
- 简单随机抽样最原始的抽样方法就是抽签摸球。该方法一般适用于总体单位比较少少的情况。

二、土地抽样调查的组织形式

- 最常用的抽样方法是利用**随机数表**——由计算机或其它随机方法制成的，即从0、1、2、……、9这10个数字出现的**概率是相**同的，但排列的先后**顺序**则是**随机**的。
- 抽取样本之前，首先应将各个总体单位**编上号**码；然后在随机数表中**任意取数**，凡是抽中的数字与相应的总体单位号码相**一致**时，该单位即为**抽中**的单位。若抽中的数字无相应的总体单位号码，则该数字被放弃，再重新抽取下一个数，直到抽满预定的样本容量 n 为止。
- 简单随机抽样从**理论**上说**最符合随机原则**，是其它抽样方式的基础，也是衡量其它抽样方式抽样效果的标准。
- **实践限制**：当总体很大时，编号困难。当总体各单位标志值之间差异很大时，采用这种抽样方式并不能保证样本的代表性。

2、机械抽样

- 又称**等距抽样**或**系统抽样**：事先将总体各单位按某一标志排列，然后依**固定顺序和间隔**抽选调查单位的一种抽样组织形式。
- 如对学生按姓氏笔划顺序排队，然后按此顺序**等间隔地抽取**样本单位进行调查。
- 等距抽样要计算**抽取间隔**，间隔 d 等于总体单位数 N 除以样本容量 n ，即 $d=N/n$ 。

2、机械抽样

- 例：从10000名职工中抽取2%即200名进行调查
- 先按姓氏笔划排列
- 按排队顺序分200组（组数等于样本容量）
- 每组50人（50也是抽取间隔）。
- 假设第一组随机抽取第5号职工，那么第一组样本单位的顺序号是5
- 第二个样本单位的顺序号是55
- 第三个样本单位的顺序号是105
- 其余类推
- 最后一个样本单位的顺序号是9955。

2、机械抽样

- 机械抽样中，由于排队所依据的标志不同，等距抽样分为按**无关标志排队**和**有关标志排队**两种。
- 1) **无关标志**：排列的**标志**和所研究的**标志值大小**无关或不起主要影响作用。例，调查职工生活水平时，职工按姓氏笔划排队；产品质量检查，按产品入库顺序排队。
- 2) **有关标志**：用来排列的**标志**和所研究的**标志值的大小**有密切**关系**。例，耕地亩产量调查，把地块按往年平均亩产的高低进行排队；职工家庭生活水平调查，把职工按工资水平高低排队。

3、类型抽样

- **类型抽样/分层抽样：**首先把总体各单位按某一标志分成若干个类型组，使**各组组内标志值比较接近**，然后分别在各组组内按随机原则**抽取样本单位**。

例：农作物产量调查，把某县60个乡按所处地貌分为三个类型区，然后从各类型区中按比例抽取9个乡作为样本。

类型区	乡数	比重（%）	各类型区抽取乡数
山区	6	10	1
丘陵	42	70	6
平原	12	20	2
合计	60	100	9

3、类型抽样

特点：把分组法和随机原则结合起来。

优点：(1) **提高**了样本的**代表性**。样本单位从各类型组中抽取，样本中有各种标志值水平的单位。

(2) **降低**了影响抽样平均误差的**总体方差**。适用于总体情况复杂，各单位之间差异较大，总体单位数又较多的情况。

4、整群抽样

- **整群抽样：**先将总体各单位划分成若干群（组），然后以群（组）为单位从总体中随机抽取一些群（组），对中选群（组）的所有单位进行全面调查的抽样组织形式。
- 例，城市居民生活水平调查，从城市全部居委会中随机抽选若干居委会，对被抽中的居委会所有住户都进行调查，这就是整群抽样。该城市的每一居委会就是一群。
- **优点：**节约和方便。整群抽样不需要编制总体单位名单，只需要编制总体群的名单。

5、多阶段抽样

- **多阶段抽样：**在抽样调查抽选样本时并不是一次直接从总体中抽取，而是分两个或两个以上的阶段来进行。
- **作用：**
- 1) 当抽样调查的面很广、总体范围太大**无法直接抽取**样本时，需要采用多阶段抽样。
- 2) 可以相对地**节约**人力和物力。从一个比较大的总体，抽取一个随机样本，势必使抽到的样本单位比较分散，若要派人调查，人力和物力的支出比较大。
- 3) 可以利用现成的行政区划、组织系统作为划分各阶段的依据，为抽样调查提供**方便**。

5、多阶段抽样

- 以某省粮食产量调查为例。可以按行政区域划分层次，以省为总体。步骤为：
 - ①从全省所有县级单位中，抽取部分县作为第一阶段抽取的样本；
 - ②从被抽中县的所有乡或村中，抽取部分乡或村作为第二阶段抽取的样本；
 - ③从被抽中乡或村的所有农户中，抽取部分农户作为第三阶段抽取的样本；
 - ④从被抽中农户的所有播种面积中抽取部分地块，进行实割实测的调查，作为最基层阶段的样本，计算其样本平均亩产量，并推算总产量。
- 多阶段抽样所划分的抽样阶段数不宜过多，一般以划分两、三个阶段，至多四个阶段为宜。

三种组织形式的区别

抽样形式	第一阶段	第二阶段
类型抽样	抽全部	抽部分
整群抽样	抽部分	抽全部
两阶段抽样	抽部分	抽部分

三、抽样指标的计算

□ （一）变量总体指标的计算

1. 总体指标公式：

简单平均数 $\bar{X} = \frac{\sum X}{N}$

加权标准差 $\sigma = \sqrt{\frac{\sum (X - \bar{X})^2 F}{\sum F}}$

加权平均数 $\bar{X} = \frac{\sum XF}{\sum F}$

方差（简单平均式） $\sigma^2 = \frac{\sum (X - \bar{X})^2}{N}$

简单标准差 $\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}$

方差（加权平均式） $\sigma^2 = \frac{\sum (X - \bar{X})^2 F}{\sum F}$

标准差也被称为标准偏差，或者实验标准差，是一组数据平均值分散程度的一种度量。标准差大，代表大部分数值和其平均值之间差异较大；标准差小，代表这些数值接近平均值。方差用来计算每一个变量（观察值）与总体均数之间的差异。为避免出现离均差总和为零，离均差平方和受样本含量的影响，统计学采用平均离均差平方和来描述变量的变异程度。

2. 样本指标公式:

简单平均数 $\bar{X} = \frac{\sum X}{n}$

加权平均数 $\bar{X} = \frac{\sum Xf}{\sum f}$

标准差 $S = \sqrt{\frac{\sum (X - \bar{X})^2}{n}}$ (资料不分组)

$S = \sqrt{\frac{\sum (X - \bar{X})^2 f}{\sum f}}$ (分组资料)

方差 $S^2 = \frac{\sum (X - \bar{X})^2}{n}$ (资料不分组)

$S^2 = \frac{\sum (X - \bar{X})^2 f}{\sum f}$ (分组资料)

例：已知某乡农民家庭月平均收入，根据资料计算总体指标及样本指标。

某乡农民家庭月平均收入资料

收入（元）	组中值（元）	农户数（户）F	抽取5%调查户f
300-400	350	120	6
400-500	450	480	24
500-600	550	100	5
合计	—	700	35

总体指标计算

收入（元）	组中值（元）	农户数（户）F	抽取5%调查户f
300-400	350	120	6
400-500	450	480	24
500-600	550	100	5
合计	-	700	35

总体平均数

$$\bar{X} = \frac{\sum X}{N} = \frac{350 \times 120 + 450 \times 480 + 550 \times 100}{700} = 447.14 \text{（元）}$$

总体方差

（加权平均式）

$$\begin{aligned}\sigma^2 &= \frac{\sum (X - \bar{X})^2 F}{\sum F} \\ &= \frac{(350 - 450)^2 \times 120 + (450 - 450)^2 \times 480 + (550 - 450)^2 \times 100}{700} \\ &= 3136 \text{（元）}\end{aligned}$$

总体标准差

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum (X - \bar{X})^2 F}{\sum F}} \\ &= \sqrt{\frac{(350 - 450)^2 \times 120 + (450 - 450)^2 \times 480 + (550 - 450)^2 \times 100}{700}} \\ &= 56 \text{（元）}\end{aligned}$$

样本指标计算：

收入（元）	组中值（元）	农户数（户）F	抽取5%调查户f
300-400	350	120	6
400-500	450	480	24
500-600	550	100	5
合计	-	700	35

$$\begin{aligned}
 \text{样本平均数 } \bar{x} &= \frac{\sum xf}{\sum f} \\
 &= \frac{350 \times 6 + 450 \times 24 + 550 \times 5}{35} \\
 &= 447.14 \quad (\text{元})
 \end{aligned}$$

$$\begin{aligned}
 \text{样本方差 } s^2 &= \frac{\sum (x - \bar{x})^2 f}{\sum f} \\
 &= \frac{(350 - 450)^2 \times 6 + (450 - 450)^2 \times 24 + (550 - 450)^2 \times 5}{35} \\
 &= 3226.9 \quad (\text{元})
 \end{aligned}$$

$$\text{样本标准差 } s = \sqrt{3226.9} = 56.806 \quad (\text{元})$$

（二）属性总体指标的计算

- 仅有两个属性标志表现的，可将属性总体分成两部分，一部分为具有某种属性标志表现的单位，另一部分为不具有某种属性标志表现的单位。
- 例：把全部产品分为合格与不合格两部分；把全部学生分为男生和女生两部分。
- 交替标志：可用“是”或“非”来表示的属性标志。为研究问题方便，常用“1”表示具有某种属性，用“0”表示不具有某种属性。
- 设总体共有 N 个单位，其中有 N_1 个单位具有某种属性， N_0 个单位不具有某种属性。则：
$$N = N_1 + N_0$$
- 总体成数

$$P = \frac{N_1}{N}$$

(二) 属性总体指标的计算

属性总体的指标计算表

交替标志 X	总体成数 F	XF	$X - \bar{X}$ ($\bar{X} = P$)	$(X - \bar{X})^2$	$(X - \bar{X})^2 F$
1	P	P	1-P	$(1-P)^2$	$(1-P)^2 \cdot P$
0	Q	0	0-P	P^2	$P^2 Q$
合计	1	P			$Q^2 P + P^2 Q$

总体平均数

$$\bar{X} = \frac{\sum XF}{\sum F} = P/1 = P$$

方差

$$\sigma^2 = \frac{\sum (X - \bar{X})^2 F}{\sum F} = P(1-P)$$

标准差

$$\sigma = \sqrt{P(1-P)}$$

- 属性总体的**样本指标**的计算方法
- 设样本 n 个单位中，有 n_1 个单位具有另一种属性， n_0 个单位具有某种属性，则样本指标为：
- 样本成数 $p=n_1/n$
- $q=n_0/n$
- 样本方差 $s^2=p(1-p)=pq$
- 样本标准差

$$s = \sqrt{p(1-p)}$$

四、抽样误差

- **统计调查误差**是统计调查结果与客观实际在数量上的差别。按产生的原因可分为登记性误差和代表性误差。
- **登记性误差**：调查过程中由于调查者或被调查者的人为因素所造成的误差。
- **代表性误差**：用样本指标来代表或推断总体指标时所产生的误差，又可分为**偏差**和**随机误差**。
- **偏差**：又称为表观误差，是指个别测定值与测定的平均值之差，用来衡量测定结果的精密度高低。
$$\sigma_{\bar{x}} = \frac{\sigma_s}{\sqrt{n}}$$
- **随机误差**：在没有登记性误差和偏差的条件下而产生的误差，是不可避免的。随机误差分为**抽样实际误差**、**抽样平均误差**和**抽样极限误差**。

（一）抽样实际误差

- **抽样实际误差**：指样本指标与用它估计未知的总体参数（总体特征值）之差。
- 具体指**样本平均数** \bar{x} 与**总体平均数** \bar{X} 的**差** $\bar{x} - \bar{X}$ ，**样本成数** p 与**总体成数** P 的**差** $(p-P)$ 。
- **总体成数**是指总体中满足某个条件的单位数与总体全部单位数的比。
- **样本成数**是指样本中满足某个条件的样本数与样本总数的比。
- 例如，某地区全部小麦平均亩产400公斤，而抽样调查得到的平均亩产为391公斤或403公斤，则样本指标与总体指标之间的误差为-9公斤或3公斤。

例：从总体1，2，3，4，5五个值中随机抽取3个数作为样本（不考虑顺序、不重复抽样），计算抽样实际误差。

□ 抽样实际误差计算表

可能样本	样本平均数 \bar{x}	抽样实际误差 ($\bar{x} - \bar{X}$)
1, 2, 3	2	-1
1, 2, 4	2.33	-0.67
1, 2, 5	2.67	-0.33
1, 3, 4	2.67	-0.33
1, 3, 5	3	0
1, 4, 5	3.33	0.33
2, 3, 4	3	0
2, 3, 5	3.33	0.33
2, 4, 5	3.67	0.67
3, 4, 5	4	1
合计	-	0

(二) 抽样平均误差

- **抽样平均误差**：所有的抽样实际误差的平方平均数。即所有可能出现的样本指标的标准差。
- 通常用 $S_{\bar{x}}$ 表示抽样平均指标的抽样平均误差。

$$S_{\bar{x}} = \sqrt{\frac{\sum (\bar{x} - \bar{X})^2}{n}}$$

- 式中：
- \bar{x} —— 抽样平均数
- \bar{X} —— 总体平均数
- n —— 样本可能数目 (或抽样平均数的个数)

抽样平均误差的应用公式

- 所有样本平均数的标准差（抽样平均误差）与总体平均数的标准差成正比，与样本容量的平方根成反比。
- 抽样平均误差的**应用公式**：

- $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

- 式中 σ ——总体平均数的标准差
- n ——样本容量
- 在**实际工作**中最常用的一种方法是用样本平均数的方差 s^2 或标准差 s 来代替 σ^2 或 σ 。

1、简单随机抽样的抽样平均误差

□ (1) 在重复抽样的条件下

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

□ s——总体标准差 n ——抽样样本数

□ 例：从20000公顷小麦中，随机抽取1000公顷实测，计算结果是平均单产量=300千克，标准差为10，计算抽样平均误差。

□ 已知：N=20000 n=1000 $\sigma = 10$ 求 $\sigma_{\bar{x}} = ?$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

$$= \frac{10}{\sqrt{1000}}$$

$$= 0.32$$

1、简单随机抽样的抽样平均误差

□ (1) 在重复抽样的条件下

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$

□ s——总体标准差 n ——抽样样本数

□ (2) 在不重复抽样条件下

$$\sigma_{\bar{x}} = \sqrt{\frac{s^2}{n} \left(1 - \frac{n}{N}\right)}$$

- 比较重复抽样与不重复抽样的误差计算公式，后者比前者多一个修正系数。在抽样中， n/N 总小于1。因此， $1 - n/N$ 总小于1。所以，不重复抽样的误差必然小于重复抽样的。同时说明不重复抽样的代表性比重复抽样的代表性高。
- 在统计工作中，通常采用不重复抽样方法进行抽样调查，但在计算抽样误差时，为简便起见，常采用重复抽样的计算公式。

2、抽样成数的抽样平均误差

□ (1) 在重复抽样的条件下

□ 其中，P为总体成数，n为样本单位数

$$\sigma_p = \sqrt{\frac{P(1 - P)}{n}}$$

□ (2) 在不重复抽样条件下

$$\sigma_p = \sqrt{\frac{P(1 - P)}{n} \left(1 - \frac{n}{N}\right)}$$

简单随机抽样的抽样平均误差计算举例

- 例：从10000件产品中随机抽取100件检查，结果发现6件是废品，计算该产品废品率的抽样平均误差。
- 已知：N=10000 n=100 p=6%求 σ_p ?

$$\sigma_p = \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.06 \times (1 - 0.06)}{100}} = 2.37\%$$

- 如果不重复抽样：

$$\sigma_p = \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0.06 \times (1 - 0.06)}{100} \left(1 - \frac{100}{10000}\right)} = 2.36\%$$

抽样平均误差的影响因素

- 第一，受总体的标志变动度 σ^2 或 σ 的大小影响。总体单位标志变动度越大，总体标准差就越大，抽样误差亦大；反之，标志变动度小，抽样误差也小。如果总体各单位标志值都相等，则标志变动应等于0，抽样误差就不存在了，抽样指标也就等于总体指标。
- 第二，受抽样单位数目多少的影响。在其他条件不变情况下，抽样的单位数越多，抽样误差越小；反之，抽样单位数目越少，抽样误差就越大。假如样本单位数等于总体单位数，抽样调查变成全面调查，抽样误差自然也就不存在了。
- 第三，受抽样方法不同的影响。重复抽样和不重复抽样计算的抽样误差是不同的，不重复抽样的抽样误差小于重复抽样的抽样误差。
- 第四，受抽样调查组织不同的影响。在同一总体中按照纯随机抽样、等距抽样、类型抽样和整群抽样四种不同调查组织形式，抽取相同容量的样本，其抽样误差是不同的。

标志变动度：综合反映总体各单位标志值变异程度的指标。简称变异指标。它显示总体中变量数值分布的离散趋势，是说明总体特征的另一个重要指标，与平均数的作用相辅相成。它可用来反映平均数代表现象一般水平的代表性程度，标志变动度愈小，则平均数的代表性愈大。可以说明现象的稳定性和均衡性。它和平均指标结合应用还可以比较不同总体标志值的相对差异程度。常用标志变动度指标有全距、四分位差、平均差、标准差等。

(三) 抽样极限误差

- **抽样极限误差**是抽样指标与总体指标之间，在一定概率保证程度下的，抽样误差的**最大可能范围**。
- 通常用 Δ 表示抽样极限误差， Δx 和 Δp 分别表示抽样平均数和抽样成数的可能误差范围。
- 抽样极限误差与总体平均数、样本平均数、总体成数、样本成数关系：
- 抽样极限误差通常需要以抽样平均误差 σ_x 或 σ_p 为标准单位来衡量，用抽样极限误差除以抽样平均误差，表示抽样极限误差为抽样平均误差的若干倍，用 t 表示，称为**概率度**。

$$\bar{x} - \Delta x \leq \bar{X} \leq \bar{x} + \Delta x$$

$$p - \Delta p \leq P \leq p + \Delta p$$

$$t = \Delta x / \sigma_x \text{ 或 } \Delta x = t \cdot \sigma_x$$

$$t = \Delta p / \sigma_p \text{ 或 } \Delta p = t \cdot \sigma_p$$

- t 称为**概率度**，其大小是由概率的大小决定的。应用时可根据要求的概率值在概率表中查出对应的 t 值，常用概率度与概率对照表如下：

t (概率度)	p (概率)	t (概率度)	p (概率)
1	0.6827	2	0.9545
1.5	0.8664	2.58	0.9876
1.64	0.9	3	0.9973
1.96	0.95		

平均数、成数的极限误差公式

平均数的极限误差公式

$$\Delta \bar{x} = t \times \sigma_{\bar{x}}$$

$$= t \times \frac{\sigma}{\sqrt{n}} \quad (\text{重复抽样})$$

$$\Delta \bar{x} = t \times \sigma_{\bar{x}}$$

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

成数的极限误差公式

$$\Delta p = t \times \sigma_p$$

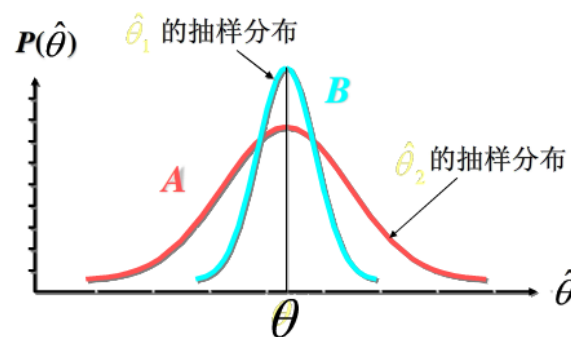
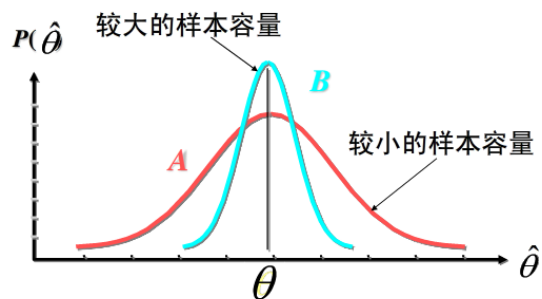
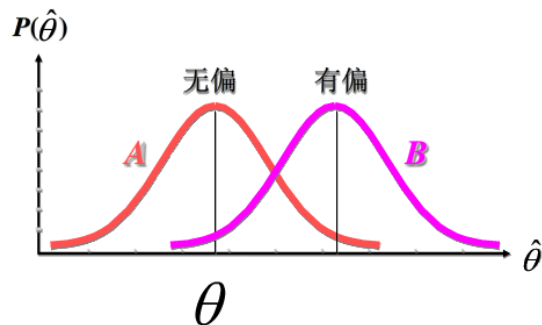
$$= t \times \sqrt{\frac{p(1-p)}{n}} \quad (\text{重复抽样})$$

$$\Delta p = t \times \sigma_p$$

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

五、总体指标估计（一）标准

- 合理抽样的基本要求：
 - **1. 无偏性。**即以抽样指标估计总体指标要求抽样指标值的平均数等于被估计的总体指标。
 - **2. 一致性。**即当样本容量 n 充分大时，若样本指标充分地靠近被估计的总体指标，则该样本指标是被估计的总体指标的**一致估计量**。
 - **3. 有效性。**即如果一个样本估计量的方差比其他估计量的方差小，则称该样本估计量是被估计的总体指标的**有效估计量**。



四、总体指标估计（二）方法

□ 1. 点估计/定值估计

- 是直接用样本指标数值代替总体指标数值，即总体平均数的点估计值就是样本平均数，总体成数的点估计值就是样本成数。这种估计不考虑是否有抽样误差。
- 例，产品合格率检测，根据抽查的样品的合格率，推断说全部产品的合格率就是该值。
- 点估计方法简单，但不很实用。因为，抽样估计中抽样指标完全等于总体指标的可能性极小。
- 样本容量多，大致推断，可用点估计

四、总体指标估计（二）方法

□ 2. 区间估计

- 区间估计所表明的是一个可能范围，不是一个绝对可靠的范围。是用样本指标和它的抽样极限误差构成的区间来估计总体指标，并以一定的概率保证总体指标将在所估计的区间内。
- 区间估计是在点估计的基础上估计出总体参数一个可能的范围，同时给出总体参数以多大的概率落在这个范围之内。

1) 总体平均数的区间估计

重复抽样条件下

$$\bar{x} - \Delta_{\bar{x}} \leq \bar{Y} \leq \bar{x} + \Delta_{\bar{x}}$$

因 $\Delta_{\bar{x}} = t \times \sigma_{\bar{x}}$ 故

$$\bar{x} - t \times \sigma_{\bar{x}} \leq \bar{Y} \leq \bar{x} + t \times \sigma_{\bar{x}}$$

$$\bar{x} - t \times \sqrt{\frac{\sigma^2}{n}} \leq \bar{Y} \leq \bar{x} + t \times \sqrt{\frac{\sigma^2}{n}}$$

不重复抽样条件下

$$\bar{x} - t \times \sigma_{\bar{x}} \leq \bar{Y} \leq \bar{x} + t \times \sigma_{\bar{x}}$$

$$\bar{x} - t \times \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \leq \bar{Y} \leq \bar{x} + t \times \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

总体平均数区间估计举例

- 某城市职工家庭调查，随机抽样400户，年平均每户消费品支出500元，标准差为100元，要求以95.45%的概率保证度，估计该城市年平均每户消费品支出。
- 已知： $n=400$, $\bar{x}=500$, $s=100$ 元, $t=2$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{100}{\sqrt{400}} = 5 \text{ (元)}$$

$$\bar{x} - t \times \sigma_{\bar{x}} \leq \bar{X} \leq \bar{x} + t \times \sigma_{\bar{x}}$$

$$500 - 2 \times 5 \leq \bar{X} \leq 500 + 2 \times 5$$

$$490 \leq X \leq 510$$

2) 总体成数的区间估计

□ 重复抽样

$$p - \Delta p \leq P \leq p + \Delta p$$

$$p - t \times \sigma_p \leq P \leq p + t \times \sigma_p$$

$$p - t \times \sqrt{\frac{p(1-p)}{n}} \leq P \leq p + t \times \sqrt{\frac{p(1-p)}{n}}$$

□ 不重复抽样

$$p - \Delta p \leq P \leq p + \Delta p$$

$$p - t \times \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} \leq P \leq p + t \times \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

- 例：某厂从60000件产品中，随机不重复抽出200件检查，发现有废品4件，要求以95%的可靠性，估计这批产品的废品率。
- 已知： $p=4/200=0.02$ ， $1-p=0.98$ ， $N=60000$ ， $n=200$ ， $t=1.96$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$$

$$= \sqrt{\frac{0.02 \times 0.98}{200} \left(1 - \frac{200}{60000}\right)}$$

$$= 0.0098$$

$$\Delta p = t \times \sigma_p = 1.96 \times 0.0098 = 0.01936 = 1.936\%$$

$$p - t \sigma_p = 2\% - 1.936\% = 0.064\%$$

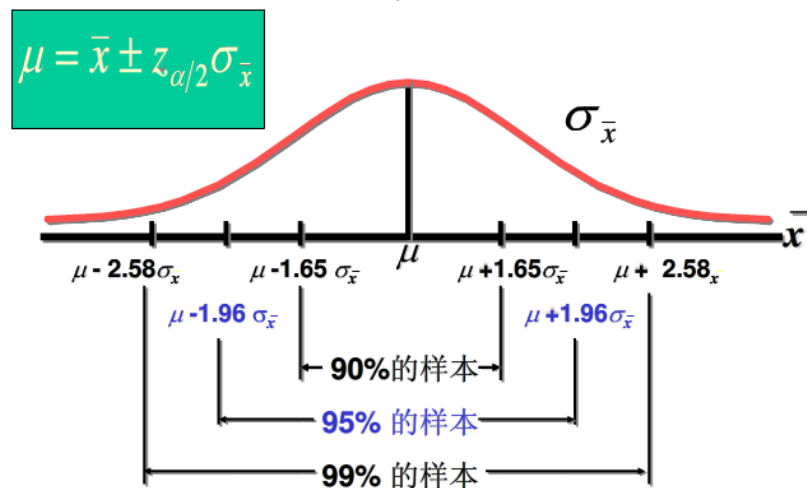
$$p + t \sigma_p = 2\% + 1.936\% = 3.936\%$$

区间估计注意问题

- 第一，用区间估计法推断的总体指标不是确定值，而是一个区间范围。
- 第二，总体指标所在的范围并不是绝对可靠的范围。这个范围总是和一定的概率相联系。
- 例如，当 $t=1$ 时，总体指标的区间范围为 $(\bar{x} - \sigma_{\bar{x}}, \bar{x} + \sigma_{\bar{x}})$ ，总体指标落在这个区间内的可能性为68.27%。
- 第三，要想提高概率把握程度，必须扩大误差范围。

$$\bar{x} - \Delta \bar{x} \leq \bar{X} \leq \bar{x} + \Delta \bar{x} \quad p - \Delta p \leq P \leq p + \Delta p$$

区间估计的图示



（三）总量指标的推算

- 1. 直接推算法
- 是指根据样本指标与另一有关的总量指标来推算所需的总量指标。
- （1）点估计推算公式
- 总量指标的估计值=样本指标值×总体单位数
- （2）区间估计推算公式

$$(\bar{x} - \Delta x) \times N \leq \bar{X} \times N \leq (\bar{x} + \Delta x) \times N$$

$$(p - \Delta p) \times N \leq P \times N \leq (p + \Delta p) \times N$$

- 例:某村种水稻2000公顷, 收割前抽查测得平均单产为500千克, 抽样平均误差为5千克, 试在95.45%置信概率下推断该村水稻的平均单产和总产的可能范围。

已知: $N=2000$, $\bar{x}=500$, $\sigma_{\bar{x}}=5$, $t=2$

$$\bar{x} - t\sigma_{\bar{x}} \leq \bar{X} \leq \bar{x} + t\sigma_{\bar{x}}$$

$$490 \times 2000 \leq \text{总产 } \bar{X} \times N \leq 510 \times 2000$$

$$500 - 2 \times 5 \leq \bar{X} \leq 500 + 2 \times 5$$

$$980000 \leq \text{总产 } \bar{X} \times N \leq 1020000$$

$$490 \leq \bar{X} \leq 510$$

2、修正系数法

- 是根据样本指标与另一有联系的指标相比较而得到的比例系数来推算或修正总量指标的方法。
- 差错数=抽样复查数-全面调查数
- 差错率=差错数/全面调查数
- 修正后数字=修正前数字 \times （1+差错率）
- 例：某县人口普查资料表明，该县总人口数为305220人，为检验普查结果的准确性，对该县5个乡抽样复查，这5个乡原先普查数字是125630人，而抽样复查结果是125785人。对该县总人口数进行修正。
 - 差错率=差错数/全面调查数
 - $= (125785 - 125630) / 305220 = 0.12338\%$
 - 修正后人数=修正前数字 \times （1+差错率）
 - $= 305220 \times (1 + 0.12338\%)$
 - $= 305597$ （人）

六、样本容量/数目的确定

- (一) 抽样单位数目确定的必要性
- (二) 抽样单位数目的计算

（一）抽样单位数目确定的必要性

- 抽样推断的基础是样本，而样本是按随机原则从全及总体中抽取一部分单位组成的集合体。
- 抽样单位数目太多，会增加抽样组织的困难，造成人力、物力的浪费；太少会使误差增大，不能有效反映总体情况，直接影响到抽样推断结果的准确性。
- 抽样推断要求推断的结果能满足在一定可靠性的条件下，保证抽样误差不超过事先规定的范围。而推断的可靠性要求主要是根据研究问题的性质和对抽样结果的用途不同而定。当可靠性要求已确定时，抽样误差的控制尤为重要。
- 抽样单位数目是影响抽样误差大小的重要因素，其它条件相同时，可以用增减抽样单位数目来控制抽样误差大小，以达到用最合适的抽样单位数满足抽样调查任务的要求。

(二) 抽样单位数目的计算

- 1、在重复抽样的条件下

- 1) 平均数的必要抽样数目

- 因为 $\Delta_x = t \sqrt{\frac{\sigma^2}{n}}$ 所以: $n = \frac{t^2 \sigma^2}{\Delta_x^2}$

- (2) 抽样成数的必要抽样数目

- $$n = \frac{t^2 P(1-P)}{\Delta_p^2}$$

2、抽样单位数目的计算

□ 2) 在不重复抽样的条件下

□ (1) 抽样平均数的抽样单位数目的计算

$$n = \frac{Nt^2\sigma^2}{t^2\sigma^2 + N\Delta_x^2}$$

□ (2) 抽样成数的抽样单位数目的计算

$$n = \frac{Nt^2P(1-P)}{N\Delta p^2 + t^2P(1-P)}$$

平均数的必要抽样数目

- 例：某县农村经济调查，已知农户平均年收入标准差为30元，要求把握程度为95.45%，允许误差为5元，试计算应抽取的样本户数。
- 已知： $\sigma = 30$, $\Delta_x = 5$, $t = 2$

$$n = \frac{t^2 \sigma^2}{\Delta_x^2}$$

- $= (2^2 \times 30^2) / 5^2 = 144$ (户)

抽样成数的必要抽样数目

- 例：对一批产品进行抽样调查，得知产品合格率为97%，要求允许误差不超过1.5%，把握程度为95%，计算需要抽取多少产品？
- 已知： $p=97\%$ ， $1-p=3\%$ ， $\Delta p=1.5\%$ ， $t=1.96$

$$n = \frac{t^2 P(1-P)}{\Delta_p^2}$$

- $= \frac{1.96^2 \times 0.97(1 - 0.97)}{0.015^2}$

- $=497$ （件）

$$n = \frac{t^2 \sigma^2}{\Delta X^2}$$

□ 3、影响抽样单位数目多少的因素

- 1) 总体中各单位的标志变异程度的大小：n与 $\delta^2 (s^2)$ 正比
- 2) 抽样推断的可靠性程度大小：n与 t^2 正比
- 3) 允许误差的大小：n与 Δx^2 反比
- 4) 抽样方法和抽样组织形式： $n_{\text{不重}} < n_{\text{重}}$

□ 4、确定抽样单位数目应注意的问题

- 1) 以上四个计算公式只适用于简单随机抽样。
- 2) 同一总体往往同时需要计算抽样平均数和抽样成数，由于它们的方差和允许误差要求不同，因此，对于抽样单位数目多少的要求也不一样，为防止抽样单位数目不足而扩大抽样误差，在实际工作中，往往根据抽样单位数目比较大的一个数目进行抽样，以满足共同要求。