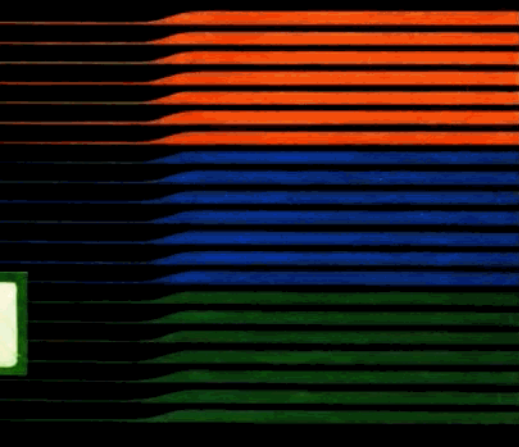


邮电通信 抽样调查

柯金水 编著



出版说明

抽样调查作为一种科学的调查方法，在社会科学和自然科学中已被广泛应用，但由于种种原因，目前这种方法在邮电通信统计中还用得不多。邮电通信由于点多面广、业务种类多、各项业务的量很大，因此，邮电通信统计很适宜采用抽样调查方法进行调查。邮电通信统计有自己的特点，在邮电通信统计中进行抽样调查，不能简单套用抽样调查原理中所介绍的各种方法。笔者试图在抽样调查理论和邮电通信统计的实践之间架设一座桥梁，把抽样调查方法引进邮电通信统计调查中来，使邮电通信统计的调查方法更加多样化、科学化。

本书在编写过程中，沈道千老师、刘立清同志曾提出许多宝贵意见，在此谨致衷心的感谢。

由于水平所限，书中难免存有不妥甚至谬误之处，诚望读者批评指正。

编 者

1988年7月

目 录

第一篇 抽样调查原理

第一章 抽样调查的基本概念	(3)
第一节 抽样调查的意义、特点及作用.....	(3)
第二节 抽样调查的理论基础.....	(7)
第三节 抽样调查的一些基本概念.....	(10)
第二章 抽样估计与必要抽样数目	(22)
第一节 抽样误差.....	(22)
第二节 总体指标的估计.....	(36)
第三节 确定必要的抽样数目.....	(43)
第三章 抽样调查的组织方式	(49)
第一节 简单随机抽样.....	(50)
第二节 分层抽样.....	(52)
第三节 等距抽样.....	(61)
第四节 整群抽样.....	(68)
第五节 两阶段抽样.....	(73)
第四章 抽样调查的工作步骤	(81)
第一节 制定抽样调查计划和抽取样本.....	(81)
第二节 资料的整理和推算.....	(83)

第二篇 抽样调查在邮电通信统计中的应用

第一章 邮电通信抽样调查	(93)
第一节 邮电业务量产品量抽样调查.....	(93)

第二节	邮电资费调查	(111)
第三节	邮电通信生产状况抽样调查	(115)
第四节	样本的轮换	(155)
第五节	市场调查	(167)
第二章	抽样检验	(177)
第一节	用抽样调查检验全面调查的质量	(177)
第二节	用抽样调查检验抽样调查的质量	(184)
第三章	各级邮电通信企业如何组织抽样调查	(193)
第一节	邮电部组织的抽样调查	(193)
第二节	由邮电部布置, 各级邮电通信企业组织 实施的抽样调查	(194)
第三节	省、地(市)邮电通信企业组织的抽样 调查	(201)
附录 1	抽样误差计算公式推导	(203)
附录 2	正态概率表	(207)
附录 3	随机数字表	(210)
附录 4	总体方差与层间方差和平均层内方差的 关系	(218)
附录 5	分层抽样的抽样误差计算公式推导	(219)
附录 6	两阶段抽样的抽样误差计算公式推导	(220)
附录 7	确定小交换机必要中继线数量的方法	(222)
附录 8	问题解答	(224)

第 一 篇

抽样调查原理



第一章 抽样调查的基本概念

抽样调查的例子在我们日常生活中随时可见，比如到集市上去买黄豆、大米等粮食，要先抓出一把看看质量如何；在家里做菜，做好之后要尝一尝，看看是咸了还是淡了；等等。在邮电统计工作的实践中，也有采用抽样调查的例子，比如平常函件日收寄量在 5 000 件以上的局，在进行平常函件件数统计时，是以平常函件的重量乘上每公斤平常函件平均件数求得的，而平均每公斤平常函件件数指标就是通过抽样调查取得的。

抽样调查是一种科学的调查方法。在我国，抽样调查方法已越来越广泛地被应用于社会生活的各方面。在农产量调查、城乡居民生活调查、物价调查、人口调查和产品质量调查等方面都已采用了抽样调查方法。相比之下，在邮电通信方面抽样调查方法应用得还比较少。随着社会经济的不断发展，各级管理部门对信息的需求量将越来越大，抽样调查作为一种科学的调查方法，必将在邮电通信方面被越来越广泛地采用。

第一节 抽样调查的意义、特点及作用

一、抽样调查的意义和特点

从全部研究对象中，随机抽取部分单位，对这部分单位

进行调查，并用部分单位的调查结果，对全部研究对象的数量特征作出具有一定准确程度和可靠程度的估计和推算，以达到对全部研究对象的认识，这种调查方法叫抽样调查。

下面举例说明。

例 1 某支局1988年9月份每天实际收寄的平常函件件数如下：

表 1-1

日 期	件数(件)	日 期	件数(件)	日 期	件数(件)
1	1 430	11	1 390	21	1 470
2	1 530	12	1 300	22	1 490
3	1 430	13	1 430	23	1 260
4	1 500	14	1 210	24	1 410
5	1 370	15	1 440	25	1 600
6	1 400	16	1 450	26	1 330
7	1 520	17	1 390	27	1 560
8	1 400	18	1 250	28	1 290
9	1 360	19	1 470	29	1 410
10	1 520	20	1 430	30	1 560

如果事先不知道每天的实际收寄量，要采用抽样调查的办法调查全月收寄的总件数，可以采用随机抽样办法，从中抽出几天进行调查。假设需要调查4天，并且采用随机方法抽中1日、9日、24日、25日，实际调查结果，这4天的收寄量分别为1 430件、1 360件、1 410件、1 600件。则这4天平均每天的收寄量为：

$$\bar{x} = \frac{\sum x}{n} = \frac{1\,430 + 1\,360 + 1\,410 + 1\,600}{4} = 1\,450 \text{ 件。这 } 1\,450 \text{ 件}$$

是 4 天的调查结果，现在要用 4 天的调查结果对九月份全月的收寄量进行推算、估计。那么该支局九月份全月收寄的平常函件的估计数为： $\sum x = \bar{x} \times N = 1\,450 \times 30 = 43\,500$ 件。这是一个抽样调查的简例。

从这个简例中可以看出，抽样调查有以下几个特点：

1. 抽样调查是一种非全面调查。同全面调查不同，抽样调查不是对研究对象的所有单位都进行调查，而仅仅是对其中的部分单位进行调查。

2. 按照随机原则从被研究对象的所有单位中抽取样本单位。这一点同典型调查、重点调查等其他非全面调查不一样。典型调查、重点调查的调查单位是有目的、有意识、有选择地抽取的，而抽样调查在抽取样本单位时是不受人的主观意志控制的，被研究对象的所有单位都有同等被抽中的可能性，各单位能否被抽中，完全是随机的。遵循随机原则，保证各单位都有同等的机会被抽中，这样就有较大的可能性使所选的样本同被研究对象的全体有相同的结构。而且，只有遵循随机原则，才有可能计算抽样误差。

3. 抽样的目的是为了推断总体。这一点也和重点调查、典型调查不同，重点调查和典型调查不能在数量上推断全体，以达到对全部研究对象的数量特征的认识。而抽样调查的目的是用样本指标去推算总体指标，用样本指标对被研究对象全体的规模、水平、结构指标作出估计。

4. 抽样调查会产生抽样误差，抽样误差可以计算出来，而且可以事先加以控制（这一点在后面讲）。

二、抽样调查的作用

目前在我国邮电通信统计方面已经有一套比较完整的报表体系和统计调查方法，那么为什么还要采用抽样调查的方法进行调查呢？

1. 在邮电通信统计方面，有些指标不可能通过全面调查取得，比如市内电话接通率指标，如果要采用全面调查，就必须在整个调查期内对所有市内电话用户都进行跟踪调查，这种调查显然是不可能的，也是没有意义的。

2. 有些指标虽然可以进行全面调查，但是要花费大量的人力和费用，而调查的目的却仅仅是为了解其变化的一般规律。如邮电业务用户构成调查、邮电业务流量流向调查等。如果对这些指标采用抽样调查，既能节约人力和费用，提高工作效率，又能达到同样的调查目的。

3. 抽样调查可以节约时间，提高调查的时效性。通信生产的时效性很强，有些指标如果采用全面调查，不但要花费大量的时间，而且可能影响通信生产的正常进行。比如平常函件业务量比较大的局，如果采用一点数的办法统计平常函件件数，就可能影响正常的封发时间。

4. 抽样调查由于调查的单位少，调查队伍经过专门培训，可以增加调查项目，取得比较详细的资料，并且可以提高资料的准确性。

5. 抽样调查和全面调查两种方法结合使用，可以发挥相互补充和检查质量的作用。

习惯上认为全面调查一定比抽样调查的质量要高，事实并非如此。有些指标采用全面调查，由于工作量大、参加调查的人员多、人员素质参差不齐、汇总的环节多等原因，出

差错的可能性就比较大，因此其调查质量可能就不够高。比如对平常函件件数进行全面调查，全国有两千多个企业，五万多个通信分支机构，每天有一千多万件平常函件，而且函件数在各企业之间的分布又非常悬殊，要一件一件地进行实际点数，数量少的单位可能不够重视，数量多的单位工作量太大，难以认真进行，因此要保证其调查结果有比较高的准确性就不大可能。如果采用抽样调查，调查单位数和调查时间就会少得多，参加调查的人员一般都经过专门训练，因此只要组织工作做得好，其调查结果的准确性一般都比较高。

当然，调查结果的准确性高低，并不仅仅取决于调查方法，采用全面调查，只要调查的组织工作做得好，调查取数认真，也是能取得很好的调查结果的。全面调查和抽样调查各有优点，有时可以结合使用，比如可以在全面调查的基础上，利用抽样调查的结果对全面调查的数字进行验证和修正。

抽样调查方法具有节约时间、节约人力、费用等优点，应用比较灵活，很适宜在邮电通信统计中推广。

第二节 抽样调查的理论基础

大数定律、正态分布和中心极限定理是抽样调查的理论基础。因此，我们在这里作一简单的介绍。

一、大数定律

大数定律又叫大数法则，它是概率论中用来阐明大量随机现象平均结果的稳定性的定理。大数定律表明，如果被研究的对象是由大量的相互独立的随机因素所构成，而且每个因素对总体的影响都相对地小，那么对这些大量因素加以综

合平均的结果，因素的个别影响将相互抵消，而显现出它们共同作用的倾向，使总体具有稳定的性质。就是说，无论个别随机因素的表现形式如何，大量随机因素的平均结果实际上是与每一个别随机因素的特征无关，并且几乎不再是随机的了。

大数定律的意义可以归纳为以下几个方面：

1. 某种现象的规律性，只有当具有这种现象的足够多单位数综合汇总在一起时，才能显示出来。

2. 某种现象的规律性，往往是以平均数的形式表现出来的。

3. 研究某种现象时，综合的单位数越多，其平均数就越能够正确地反映出该现象的规律性。

4. 大量现象的某种数量标志密集在平均数附近，这些现象共同决定着平均数的水平；个别偏离平均数比较远的因素，它们对平均数的影响则会由于足够多数单位综合汇总的结果，而相互抵消，趋于消失。

大数定律证明，当抽样单位数 n 足够大时，可以有接近100%的概率，期望样本平均数与总体平均数的绝对离差为任意小。这就从理论上揭示了样本和总体之间的内在联系，即随着抽样单位数 n 的增加，样本平均数会越来越接近于总体平均数。大数定律所论证的样本平均数趋近于总体平均数的趋势，为抽样推断、估计提供了重要的理论依据。

二、正态分布

许多社会现象中，虽然在各个单位之间某个标志的数值大小不等，但是通过大量的观察可以发现，大多数单位的标志值接近总体平均数，标志值同总体平均数的绝对离差比较

大的单位数比较少，标志值越偏离总体平均数，这样的单位数就越少。而且，标志值高于总体平均数的单位数同标志值低于总体平均数的单位数基本上是对称的。总体现象的这种分布称为正态分布。正态分布有如下几个特点：以总体平均数为中心，高于总体平均数的单位数和低于总体平均数的单位数基本相等；而且正离差和负离差相等的单位数也基本相等；接近总体平均数的单位数占大部分，远离总体平均数的单位数所占比重很小，而且逐渐趋于零。

比如，我国成年男同志的平均身高约为1.7米，大部分男同志的身高在1.6米至1.8米之间，低于1.6米和高于1.8米的人所占的比重比较小，特别矮和特别高的人很少，身高在1.4米以下或2米以上的人所占的比重接近于零。在1.65米左右和在1.75米左右的人数基本相等，在1.6米左右和在1.8米左右的人数也基本相等。

在呈正态分布的总体中抽样，因为大部分单位的标志值接近总体平均水平，因而抽到这样的单位的可能性就比较大，这样就能有较大的把握保证样本平均数接近总体平均数，这就为抽样调查提供了又一理论依据。

三、中心极限定理

中心极限定理是研究变量和的分布序列的极限的原理。实际上，并非所有的社会经济现象都是服从正态分布的。在非正态分布的总体中抽样，能否保证样本平均数对总体平均数有足够的代表性呢？中心极限定理证明了，不管总体变量的分布如何，随着抽样单位数 n 的增加，抽样平均数的分布便趋近于正态分布。变量和分布在邮电通信中是普遍存在的，如某邮电局每天收寄包裹总重量，是该局每天收寄的每件

包裹重量之和,某邮电局每天投递的邮件总重量,是该局每天投递的每件邮件重量之和;某报房每天拍发的电报总字数,是该报房每天拍发的每份电报字数之和;某长话台每天处理的去话总时长,是该台每天处理的每张去话时长之和;等等。根据中心极限定理,这些独立因素之和的分布都是趋近于正态分布的。样本平均数也是一种随机变量和,因此在抽样单位数 n 足够大的情况下,样本平均数的分布也趋近于正态分布。中心极限定理也是抽样推断的一个重要的理论依据。

第三节 抽样调查的一些基本概念

一、全及总体和总体单位

1. 全及总体。全及总体就是调查对象的全体,简称总体。总体由具有某种共同性质的许多单位组成,是具有同一性质的许多单位的集合体。如例1,要调查9月份平均每天收寄的平常函件件数,则9月份30天便构成了总体。又如,要调查某企业职工平均工资,则该企业的职工便构成了这个调查的总体。总体是我们最终要研究的对象。

2. 总体单位。构成总体的基本单位叫总体单位。如上述9月份30天构成了一个总体,则每一天便是总体单位;企业的所有职工构成了一个总体,则每一个职工便是总体单位。总体单位数一般用“ N ”表示,在例1中, $N=30$ 。

在同一总体中,总体单位的样式不是一成不变的,它随着调查目的、调查方法、调查的客观条件等的变化而改变。在例1中调查每天的收寄件数以推算全月的收寄总件数,这时总体单位是“日”,总体单位总数是30。如果是要调查总件

数中各类用户交寄的件数所占的比重，即要调查用户构成，这时的总体单位是“件”而不是“日”，总体单位总数是总件数而不是总天数。当总体单位改变时，一些总体指标和样本指标的计算都要跟着改变。

二、样本和样本单位

样本是相对于总体而言的，是从总体中根据随机原则抽出来进行观察的一部分总体单位。样本单位就是组成样本的各个总体单位。样本单位数称为样本容量，一般用 n 表示。相对于总体单位总数 N 来说，样本单位数 n 是个很小的数。一般说，样本单位数达到或超过 30 个便称为大样本，不足 30 个的称为小样本。在邮电通信专业抽样调查中一般都是大样本。

样本单位是总体单位，但总体单位不一定是样本单位。在抽样过程中，总体单位被抽中了，总体单位同时也就是样本单位；如果没有被抽中，便不是样本单位。

三、总体指标和样本指标

总体指标。根据全部总体单位计算的统计指标叫总体指标。在例 1 中，总体单位总数是 30，根据 30 天收寄的平常函件件数计算的全月平均每天收寄数就是总体指标。在一个总体中，总体指标是一个唯一确定的、客观存在的量，但在抽样调查所研究的总体中，总体指标是个未知数，抽样调查的目的，就是要对总体指标进行推算、估计。

样本指标。根据全部样本单位计算的统计指标叫样本指标。在例 1 中，从总体 30 天中抽出 4 天组成样本，根据这 4 天收寄的平常函件数计算的平均每天收寄数就是样本指标。在一个总体中，有很多可能样本，每个样本都有一个样

本指标。抽样调查是从许多可能样本中随机抽取一个样本。由于样本是随机抽取的，不知道会抽到哪个样本，不同的样本就有不同的样本指标，因此，样本指标的值是一个随机变量。

抽样调查过程中所研究的总体指标和样本指标主要有：平均数、成数、方差和标准差。

平均数。平均数是反映总体一般水平的指标。总体中各单位某一数量标志的数值有的高有的低，有的大的有的小，把它们综合起来后，求出平均水平，作为总体的代表值，这就是平均指标的主要作用。抽样调查的一个主要目的，就是要求出样本平均数，以推断总体的一般水平。在例 1 中，每天收寄的平常函件数各不相同，从中随机抽出 4 天，综合之后求出样本平均数，以作为对总体进行估计的主要依据。

以 N 表示总体单位总数， n 表示样本单位总数。 x 表示总体某个标志的变量，则总体平均数的计算公式为：

$$\bar{X} = \frac{\sum_{i=1}^N x_i}{N}, \text{ 用文字表达为：总体平均数} = \frac{\text{总体标志总量}}{\text{总体单位总数}};$$

$$\text{样本平均数的计算公式为：} \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \text{ 用文字表达为：样}$$

$$\text{本平均数} = \frac{\text{样本标志总量}}{\text{样本单位总数}}.$$

成数 成数指标也叫比重指标，它是一种结构相对数，是指在总体中具有某种属性的总体单位数占总体单位总数的比重。比如某企业有职工 1 000 人，其中女职工 450 人，占总数的 45 %，这就是成数指标。又如某局对市话接通情况

进行测试，拨叫 500 次，其中接通 360 次，接通次数占拨叫总次数的 72%，这也是成数指标。成数也可以看作是平均数，它是一种特殊的平均数。在属性总体中，各单位的标志不能用数量表示，只能用一定的术语加以描述。为了反映属性总体的数量特征，我们把描述属性总体单位标志的术语转换成数量，即凡具有所要调查的标志的单位，其标志值用“1”表示，凡不具有所要调查标志的单位，其标志值用“0”表示。这样就可以用计算平均数的办法计算成数指标。在上面的例子中，是女职工的，标志值为 1，不是的为 0，则女职工占职工总数的比重 = $\frac{\text{总体标志总量}}{\text{总体单位总数}} = \frac{1 \times 450 + 0 \times 550}{1\,000} =$

45%。在市话接通率调查中，是接通的，标志值为 1，不是接通的，标志值为 0，则接通率 = $\frac{\text{总体标志总量}}{\text{总体单位总数}} = \frac{\text{接通总次数}}{\text{拨叫总次数}} = \frac{1 \times 360 + 0 \times 140}{360 + 140} = 72\%$ 。

总体成数指标一般用“ P ”表示。如果以 N 表示总体单位总数， N_1 表示总体中具有所要调查标志（简称“是”）的单位数， N_0 表示总体中不具有所要调查标志（简称“非”）的单位数，则 $P = \frac{N_1}{N}$ ， $Q = \frac{N_0}{N}$ 。因为 $N = N_1 + N_0$ ，所以 $P + Q = \frac{N_1}{N} + \frac{N_0}{N} = 1$ ， $P = 1 - Q$ 。属性总体单位只有“是”和“非”两种表现形式。样本成数指标一般用“ p ”表示，如果以 n 表示样本单位总数， n_1 表示样本中属于“是”的单位数， n_0 表示总体中属于“非”的单位数，则 $p = \frac{n_1}{n}$ ， $q = \frac{n_0}{n}$ ， $p = 1 - q$ 。

成数指标和平均指标的调查。 当总体单位的样式变化