

## 第十四章 社会统计资料的搜集与整理

社会统计资料的搜集与整理指人们在一定的理论与方法指导下,从客观实际出发,在实际生活中系统地、直接地搜集各种具体的原始材料,对其进行整理和加工的过程。它是人们认识社会与了解社会的基础。通过它,人们可以从感性认识入手来了解社会的真实情况;把握社会现象的一般状况、过程和特点;解释社会现象的产生、发展和变化,揭示社会现象的本质和发展规律,并可预测与控制社会现象的发展规律,为我们制定有效的政策和措施及解决社会问题等提供科学的依据。

### 第一节 社会统计资料的调查方式

社会统计资料的调查方式可以以多种角度,按不同的标准进行分类。<sup>①</sup>依据调查对象的范围及运作方式,可将其划分为统计报表、普查、抽样调查、典型调查、重点调查和个案调查六种类型。

#### 一、统计报表

##### 1. 统计报表的特色

统计报表是有计划、有组织地取得统计资料的一种调查组织形式,是一种经常性的全面调查方法。统计报表由国家和代表国家行使管理职能的上级主管机关统一制定颁发,逐级下达,要求基层单位以一定的原始记录为基础,按照统一规定的表格形式、指标内

---

<sup>①</sup>见袁方《社会调查原理与方法》第二章,高等教育出版社

容、统计范围、报送时间和报送程序逐级汇总上报,在中国,这种资料的搜集已成一种制度。迄今为止,中国官方出版印行的各种统计年鉴中的大部分统计资料,都是以这种方式获得的。统计报表这种调查方式具有以下特点:

(1)具有统一规定的表格、填写方法和报送期限,保证所提供的资料具有统一性和时效性。

(2)以原始记录为依据,所提供的资料必须数出有据。

(3)对各项指标规定了严格的统计内容和计算方法。

(4)可以连续、完善地积累资料。

(5)是各级党政领导部门指导工作、制定政策,编制、检查和监督计划的主要依据。

当然,以统计报表获取资料,也存在一定的问题。如指标设计的重复与不衔接,表格内容固定,缺少灵活性,难以反映社会经济发展的新情况,且经常大量地占用人力与物力,因此,它并不是我们获取资料的唯一方法。

## 2. 统计报表的种类

统计报表可以按不同的标准而区分成多种类型。

(1)按主管系统和报表内容的差别,统计报表可区分为国家统计报表、部门统计报表和地方统计报表。

(2)按报送周期长短的不同,统计报表可分为日报、旬报、月报、季报、半年报和年报。除年报外,其他报表均称为定期报表。

(3)按填报单位的不同,统计报表可分为基层统计报表和综合统计报表。

(4)按报送的方式不同,统计报表可分为邮寄报表和电讯报表(包括电报、电话和电视传真等)。

## 3. 统计报表的内容

统计报表的基本内容包括报表目录、表式、填表说明和统计目录四个部分。

报表目录是指说明应报送的各种报表的名称、填报范围、报送

日期及有关事项的一览表。报表目录主要包括以下具体项目：表号；报表名称；报告期别；统计范围；报送日期；受表机关。

表式是指统计表的具体格式。各种表式是按一定规格制订的，表内主要包括要求填报的指标项目及表外填报的各项补充资料。

填表说明主要包括填报范围和指标解释两部分。

统计目录是指应在有关报表的主栏中填报的统计分组及具体项目的一览表。

## 二、普查

### 1. 普查的特点

普查也称整体调查或全面调查，它是为了了解研究对象的一般情况而对较大范围的地区或部门中的每个对象都无一例外地进行调查的方法。统计部门和政府有关部门为了了解基本国情，把握社会总体的全貌，得出具有普遍意义的结论，经常进行各种普查，如人口普查、工业普查、农业普查等。普查具有以下特点：

(1)它是专门组织的一次性全面调查，是一种不连续的调查方法，其调查对象具有时点性。

(2)组织领导的高度集中和统一。无论在一国或一个局部地区进行的普查，均应统一筹划和部署，统一内容、时间和调查方法，以保证资料的可靠性与准确性。

(3)调查时间的标准性与特定性。普查所观察的是某一时点状态下的定量和定性特征，获得的资料是静态的，故普查时必须确定一个与调查对象的正常状态大体一致的标准时点。如人口现象的出生、死亡、移动一般在6、7月份相对稳定，变动较小，故宜在此期间内选择标准时点。此外，为保证资料在静态上的对比和动态上的连续性，标准时点一经确定，不能轻易改变。

(4)普查的结论具有很高的概括性和普遍性，可以精确地反映社会总体的一般特征。

(5)普查涉及的面广，工作量大，时间性强，对调查工作本身要

求甚高,并需耗费大量的人力与物力。以中国 1982 年第三次人口普查为例,这次普查从准备到正式结束共历时约 6 年,动员了 500 多万人作调查员,花费资金 4 亿元。

(6)调查的内容具有局限性。普查不可能对每一个调查对象都进行深入细致的调查。故普查调查的项目较少,资料缺乏深度。

## 2. 普查的组织

普查的组织方式主要有两种:一种是自上而下地组织专门的普查机构和调查人员,直接对调查单位进行登记调查。如人口普查。另一种是不专门设立统一的普查机构和配备调查人员,而是利用企、事业单位的组织系统及其平时的核算资料和行政记录进行全面调查。如对科技人员数量、学历等的普查。

普查涉及的范围广,工作量大,一次普查需要动用许多人力与物力,故必须进行认真的组织与筹划。普查工作大体可分三个阶段:

(1)准备阶段。包括建立统一的组织领导机构,设计普查方案,确定普查对象和单位,普查项目和普查时间;配备和训练参加普查的各种工作人员;物资和技术的准备及在小规模范围内进行试点。

(2)组织实施阶段,又称正式普查、全面搜集资料阶段,这是普查工作的高峰,要在限定的较短时间里,集中人力与物力,全力以赴地完成全部的调查任务。

(3)复查、审核、汇总资料并撰写有关报告。

由于普查所需的时间、人力、经费都很多,因此,除统计部门和政府部门外,一般社会统计资料的获取很少采用普查。

## 三、抽样调查

### 1. 抽样调查的概念与特点

抽样调查是从调查对象的总体中抽取某些单位(或个人)作为样本进行调查,并以样本的状况来推论总体状况的调查类型。从总体中取出部分的过程叫抽样,抽出的这一部分单位称作样本。抽样

调查的目的是从许多“点”的情况来概括“面”的情况。成功的抽样调查具有：

(1)调查费用较低,因为抽样调查涉及的只是总体中的一部分。

(2)速度快,能够及时了解和把握社会现象的变动。

(3)应用范围广。抽样调查可用于各个领域,各个部门,各个课题,而不象普查那样只局限于某些部门。

(4)可获得内容丰富的资料。抽样调查可设置较多和较复杂的调查项目,并能集中时间和精力作详细的分析。

(5)资料的准确性高。抽样调查可以使用少量素质较高的工作人员并对他们进行充分的训练,还可以在实地调查中给予更仔细的检查与监督,调查资料的处理亦能较好完成,因此,与普查相比,抽样调查的资料往往更准确、可靠。

抽样调查成功的关键在于所选取的样本能够代表总体,即抽取出来的样本从调查所要研究的总体特征看,能够再现总体的结构。样本的代表性主要取决于以下三个因素。

首先是抽样方法的选择。抽样方法可分为概率抽样与非概率抽样两种。概率抽样依据抽样理论和严格的抽样程序,使总体中每个单位被抽取的概率为已知。非概率抽样只是根据研究任务和对调查对象的分析,主观选取样本。用概率抽样法抽取的样本能有较高的代表性,因为它避免了抽样者在抽样过程中有意或无意的偏差(也称系统性误差)。

其次是对总体的界定与了解。总体是在理论上明确定义的研究单位的集合体。例如关于妇女生育率的研究中,15-49岁有生育能力的妇女即为调查研究的总体。严格的概率抽样可获得代表性较高的样本,但这是建立在对总体正确界定和了解的基础上。否则,尽管采用随机抽样方法,仍难免失败。如1936年美国《文学摘要》关于总统选举民意测验的失败,就是因为对总体的错误界定,它以电话簿和汽车注册簿上的选民为总体、忽略了没有家庭电话

和私人汽车的选民。

第三是样本的大小,又称样本容量,指样本内所含单位的数量。样本的代表性与样本的大小有密切的关系。样本越小,代表性越低,但样本越大,抽样的成本就越高。样本数的确定应考虑容许误差的大小、总体的性质及客观的人、财、物的制约。

## 2. 抽样的基本程序

抽样调查的主要程序是:

(1) 界定研究总体和调查总体;

(2) 编制总体名单,即编制业已确定的抽样范畴内的抽样单位名单。如进行家庭调查,可先在一个县中抽取若干个乡,然后在由这些乡组成的样本中抽取某些村,最后从这些村组成的样本中抽取家庭,则应编制的总体名单为:全部乡的名单,乡样本中所有村的名单,村样本中所有家庭的名单。

(3) 确定样本大小和选样抽样方法;

(4) 抽取样本;

(5) 搜集资料;

(6) 对样本的代表性进行评估。

## 3. 概率抽样

概率抽样可以分为简单随机抽样、等距抽样、分层抽样、整群抽样以及结合使用上述两种或两种以上抽样方法的多阶段抽样。

### (1) 简单随机抽样

简单随机抽样是最基本的概率抽样,分为重复抽样和不重复抽样。重复抽样即每次抽中的单位仍放回总体,样本中的单位可能不止一次被抽中。不重复抽样即抽中的单位不再放回总体,样本中的单位只能抽中一次。社会调查往往采用不重复抽样。

简单随机抽样的具体作法有两种,一是抽签法,即将总体中的全部单位逐一作签,搅拌均匀后进行抽取。二是随机数字表法(见附表1)即将总体中所有单位编号,然后在随机数字表中确定一个随机起点进行抽取。

简单随机抽样必须有一个完整的总体各单位的清单,当总体太大时,由于制作清单很困难,故这种抽样方法很少采用。

### (2)等距抽样

又称系统抽样,它是简单随机抽样的变种。进行等距抽样时,应先将总体从1~N依次编号,并计算抽样距离 $K=N/n$ (式中N为总体单位总数,n为样本容量),然后在1~K中抽一随机数 $k_1$ 作为样本的第一个单位,接着取 $k_1+K, k_1+2K, \dots$ 直到抽够n个。如设某大学有10 000名学生,决定采用等距抽样的方法从中抽查200名学生,这时的等距离 $K=\frac{10\ 000}{200}=50$ ,用简单随机抽样的方法在前50名学生中抽出第一位学生,设其序号为8,则第二位学生的序号为 $8+50$ ,第三位学生的序号为 $8+2\times 50, \dots$ 即每隔50人抽1人,直至抽够200人。

等距抽样在运作时要防止周期性偏差,因为它会降低样本的代表性。如军队人员名单通常按班排列,10人一班,班长排第1名,若抽样距离也取10,则样本全由士兵组成或全由班长组成。

### (3)分层抽样

分层抽样指先依据一种或几种特征将总体分为若干个子总体,每一子总体称作一个层,然后从每层中随机抽取一个子样本,这些子样本合起来就是总体的样本。常用的确定各层样本数的方法有两种:一是分层定比法,即各层样本数与该层总体数的比值相等。例如,总体 $N=500$ ,样本大小 $n=50$ ,则样本比例为 $n/N=0.1$ ,每层均按这个比例确定该层样本数。二是非比例分配法,当某个层次包含的个案数在总体中所占比例很大时,为使该层的特征在样本中得到足够的反映,可人为地适当增加该层样本数在总体样本中的比例。当然,这样做会增加推论的复杂性。

### (4)整群抽样

又称聚类抽样。先将总体按照某种标准分为不同的群,每个群为一个抽样单位,用随机的方法从中抽取若干群,对抽中的样本群

中的所有单位进行调查。整群抽样分类的原则是使群与群之间的差别尽可能小,而群内部的差别尽可能大。

#### (5)多阶段抽样

多阶段抽样是上述四种抽样方法中的两种或数种的综合。下面是一个某市妇女生育力调查的抽样过程,这一调查欲抽取1 000名妇女以了解全市妇女的生育情况。调查的总体是育龄(15-44岁)妇女中已婚且至少有一个小孩、目前仍与丈夫住在一起的人。具体抽样过程如下:

①此市有东、中、西三个经济区,可作为三个层,实施分层抽样。为此先计算每个经济区抽取人数比例(人数=户数)。即:

$$X(\text{每个经济区抽取人数比例}) = \frac{\text{经济区内总户数}}{\text{三个经济区内总户数}}$$

$$Y(\text{每个经济区抽取人数}) = 1\,000 \times x$$

假定东、中、西各区分别为5万户、3万户、2万户,则 $x$ 分别为0.5,0.3,0.2; $Y$ 分别为500人,300人,200人。

②用整群抽样从三个经济区内抽取10个街道,则:

$$Z(\text{每个经济区入选街道}) = \frac{\text{本经济区内街道总数}}{\text{三个经济区内总街道数}} \times 10$$

假定东、中、西三区分别有8个、8个、4个街道,则 $Z$ 分别为4个、4个、2个。

③用整群抽样法从每个街道抽取3个居委会,则 $V$ (每个经济区抽取的居委会数) $=3 \times Z$ 。东、中、西三区中应抽的居委会数分别为12个,12个,6个。

④计算每个居委会应抽妇女数,即:

$$W(\text{每居委会抽取的妇女数}) = Y/V$$

则东区: $W_1 = 500 \text{ 人} / 12 = 42 \text{ 人}$ ,中区 $W_2 = 300 \text{ 人} / 12 = 25 \text{ 人}$ ;西区: $W_3 = 200 \text{ 人} / 6 = 33 \text{ 人}$ 。

最后,在所抽出的居委中依年龄组:15-19岁,20-24岁,25-29岁,30-34岁,35-39岁,40-44岁均匀地随机抽选妇女。

#### 4. 非概率抽样



根据研究任务的要求和调查的性质,主观地、有意识地从总体中选取被认为在某种意义上能较好代表总体样本的方法。非概率抽样分为:

#### (1) 偶遇抽样

研究者将在某一时间和环境中所遇到的每一总体单位均作为样本成员,如街头拦人法。

#### (2) 立意抽样

又称判断抽样。研究人员从总体中选择那些被判断为最能代表总体的单位作样本的抽样方法。常用于总体小、内容差异大,且总体边界无法确定的情况。

#### (3) 定额抽样

又称配额抽样,它是按调查对象的某种属性或特性将总体中所有个体分成若干类或层,然后按各层样本数与该层总体数成比例的原则主观抽取样本。如研究家务劳动的分配,考虑到婚姻与性别都可能对其有影响,可将研究对象分为未婚男性、已婚男性、未婚女性和已婚女性四个组,然后从各个组非随机地抽样。

#### (4) 滚雪球抽样

以若干个具有所需特征的人为最初的调查对象,然后依靠他们提供认识的合格的调查对象,……依次类推,样本如同滚雪球般由小变大。如对妓女、同性恋、吸毒等的研究,常用这种方法。

### 四、典型调查与重点调查

#### 1. 典型调查

典型调查是从研究总体中有意识地挑选一个或几个有代表性的单位进行全面、深入的调查,以达到了解总体特征和本质的方法。这里的单位可以指企业、学校,也可指个人、家庭、村庄、城市或班组、团体等。典型调查的目的是通过深入地“解剖麻雀”,以少量典型来概括或反映全局。实施典型调查的主要步骤是:

(1) 根据研究目的,通过多种途径了解研究对象的总体情况。

(2)从总体中选出备选单位,加以比较,慎重选出有较大代表性的典型。

(3)进点调查,具体搜集材料。

(4)分析研究材料,得出结论。

典型调查涉及的单位少、省时、省力、省钱,可以对研究对象进行深入细致的调查,但它无法用概率原理推断总体的数值及可靠程度,且常会受研究者主观作用的影响。

## 2. 重点调查

重点调查是指在所要调查的对象中,只选样调查一部分重点单位。所谓重点单位,一是指在全局中处于举足轻重地位的某些地区、部门、单位、事项;二是指在工作方法中应该抓住重点,以便由点到面,开展工作;三是指作为重点调查中的重点,主要是指它们的标志量在总体标志总量中占有很大比重的单位。重点调查常用于统计工作和行政管理中。如在全国几十个大中型钢铁企业中选取首钢、宝钢、鞍钢等 10 个重点企业进行调查,就能大致了解全国钢铁生产的情况,因为这 10 个企业的钢产量占全国钢产量的 50% 以上。

## 五、个案调查

个案调查是从总体范围内选取个别研究对象进行调查的方法。通常选取比较有典型意义的对象来进行调查。个案调查的主要目的不是要由个体推论总体,而是要深入、细致地描述一个具体单位(如个人、家庭、群体、企业、社区等)的全貌和具体的社会过程。个案调查作为一种有效的研究方法,在 19 世纪和 20 世纪初期曾是社会调查研究的主要方式。例如,研究人员从工人、农民、贫民、乞丐、娼妓、原始部落、村庄、社区、企业等单位中选取一个或几个调查对象作为个案,详细、深入地了解每一个调查对象的社会活动、生活方式、行为模式、价值观念、文化、规范等,通过对个案的详细描述与分析,可以发现重要的社会现象与规律,为社会研究提供

有用的范畴。由于个案调查可以对对象做深入的质的研究,彻底把握其全貌,因而它具有抽样调查所不具备的社会实在性。

进行个案调查的方法一般有参与观察、深度访问、重点访问、生活史研究、个人文献研究、社区研究等几种。

## 第二节 社会统计资料的搜集方法

对社会状况、社会问题以及社会发展趋势的了解、分析与把握,是建立在能搜集到真实、准备的资料的基础上的。搜集资料的方法有许多种,如定性研究搜集资料的方法有文献法、访问法和观察法;定量研究搜集资料的方法有问卷法、量表法和测验法。每一种方法都有自身特定的意义和作用。本节主要介绍观察法、访问法、问卷法和量表法这四种常用的方法。

### 一、观察法

观察法是一种主要依靠研究者的感官(视觉、听觉等)或借助科学仪器来搜集资料的方法。它与日常生活中的观察有所不同,它根据研究者的研究目的来选择观察对象,在观察以前对观察对象的情况已有一定的了解,对于观察的内容、方法,进入观察场所和与被观察者相处的方式等,都有一定的计划。

依照观察者参与到观察对象和环境活动中的程度,观察法可分为参与观察和非参与观察两种形式;依照观察者对观察环境进行人工设计的程度,观察法可分为有结构观察和无结构观察两种。

#### 1. 参与观察

参与观察又可分为完全参与观察与半参与观察。

完全参与观察指观察者隐瞒自己的调查目的和真实身份,完全以普通一员的身份参加到被观察的群体中去,和他们一样正常活动,进行悄悄观察。这种方法有很大的灵活性,能作深入的探索并获得深层和隐秘的资料。但要求观察者能用比较长的时间和被

观察者生活在一起,完全投入其中,被群体认同,并在不引起被观察者注意的情况下,作好详细、准确的记录。

半参与观察指观察者虽置身于被观察的环境和活动中,但并不与他们融为一体,而是有选择地参加他们的活动,并始终保持观察者的身份。这种方法因观察者的特殊身份,不是群体的一员,因此,一些较隐秘的事情就难以了解到。

## 2. 非参与观察

非参与观察指观察者处于旁观的地位,不参与被调查者的任何活动。这种方法要求观察者处于不为观察对象察觉的位置,以便搜集到最为自然、客观的资料。

## 3. 有结构观察

有结构观察是按照预定的标准化的观察项目和程序进行的观察。有结构观察的程序是:

(1)确定观察对象和观察内容。观察对象是依据研究的任务和目的设定的。观察内容一般包括以下五个方面。一是环境,它是被观察对象活动的舞台,可对被观察者的行为产生很大的影响。被观察对象所处的环境是怎样的?对于社会行为鼓励什么,反对什么?环境发生了什么变化,怎样变化的?二是被观察者。人是社会行为的主体,故是观察的焦点。被观察对象中有什么样的人?相互间关系如何,有何特点?关系是如何形成的?三是被观察对象的社会行为。他们在做什么?如何引起的?造成了什么后果?发展趋势如何?四是被观察者的价值判断与态度取向。他们赞成什么,反对什么?五是频率与持续期。现象何时发生?出现的频率与持续时间如何?

(2)将观察的内容具体化,作出详细的分类,确定观察的变量与指标。

(3)依据观察指标设计观察表格、卡片或拟定观察提纲,并规定标准化的观察方法和记录方法,然后据此实施。

## 4. 无结构观察

无结构观察是观察者并不预先规定标准化的观察项目和程序,仅根据研究目的和任务灵活进行的观察。无结构观察通常分为无结构非参与观察和无结构参与观察。

## 二、访问法

访问法是访问者通过口头交谈的方式向被访问者了解社会情况的方法,其最大特点是通过交谈获取资料。访问法可作为搜集资料的主要方法,也可作为辅助方法去验证或补充其他方法获得的资料。

### 1. 访问法的类型

访问法可根据不同的标准进行分类,按探究类型可分为结构性访问与非结构性访问;按接触期间和时间可分为一次性访问和追踪访问;按参与人数可分为集体式访问和个别访问。

#### (1) 结构性访问与非结构性访问

结构性访问是访问者在访问前就制定好了详细的标准化的访谈提纲。进行结构性访问时,所有的访问员都要遵循事先制定好的访谈提纲,按照一定的顺序提出问题,不能随意改变访谈内容。访谈提纲的制定是进行结构性访问的关键,研究人员往往将需要询问的问题制成统一的访问调查表。常用的访问调查表是问卷。

非结构性访问是访问者不依照某种统一的访问调查表,而是围绕研究的问题与被访者进行自由交谈,访问对象可以随便提出自己的意见。无结构性访问依访问的目的、内容和实施方法的不同,可分为重点访问、深度访问与客观陈述三种。重点访问又称集中访问,常用于对广播、电视、出版物等的社会及心理效果的研究。如让被访者参加心理测验并让其说出感受或认识。深度访问又称临床式访问,常用于个案调查。客观陈述法又称非引导式访问,被访问者常常在访问者的鼓励、询问下,把自己的信仰、价值观、行为以及其所生活的社会环境进行客观描述。在这种类型的访谈中,经常的提问是:您认为某事怎么样?为什么这样?

## (2) 一次性访问与追踪访问

一次性访问是对人们某一生活时刻或某一时期内的思想、态度、行为等情况而进行的调查,常用于对某个特殊问题或某些事件发生后人们的态度、行为等的了解与描述。

追踪访问是通过调查人们随时间的推移在不同的生活时期其思想、态度以及行为的变化,它属于深度访问的一种,具有较高的科学研究性质。

## (3) 集体式访问与个别访问

集体式访问是通过一名或几名访问员亲自召集一些人开座谈会以搜集资料的方法。开座谈会搜集资料应注意:第一,参加人员要有代表性;第二,参加的人数不宜太多,根据长期的实践经验,座谈会的人数以5~7人为宜;第三,参加人员要十分了解情况并敢于发言;第四,要注意造成座谈的气氛,故在开始时宜说些轻松的话题;第五,要注意控制座谈会的主题;第六,尽量造成民主、平等的讨论气氛。

个别访问是由访问员对被调查对象逐一进行面对面的询问,并将回答记录下来方法。个别访问的成功与否取决于:第一,该问的问题一定不能遗漏;第二,对所问的问题一定要问清楚明白;第三,对询问过的问题做真实的记录;第四,访问员的人际交往技巧;第五,受访人了解情况的多少、合作意愿及言语表达能力等。

### 2. 访问的程序

访问的进行一般可分为五个环节,即访问的准备,调查员进入访问,与被调查者建立合作关系,正式访谈及结束访问。

#### (1) 访问的准备

访问的准备工作包括根据研究的目的选择适当的访问方法,制作相应的问卷,调查表或调查大纲,了解被访人所在社区的社会文化环境和行为特征,对要作的工作与时间进行安排,以及对必要的调查所需工具(如笔、纸、调查表格、说明书、问卷、照相机等)的准备。

## (2) 进入访问

在实地访问前,先要与调查对象所属的省、市、区、乡政府机关或派出所等地方机构取得联系,获得对方的允许与支持,然后是正面接近被访问者。这时,访问员必须进行自我介绍(必要时出示有关介绍信和身份证),说明来访的目的,为什么要进行这项研究,并认真地保证对被访人的言行保密,以争取他的支持与合作。此外,还应向被访人解释他是如何被选出来的,以及其对问题的回答对研究的意义。这一阶段最易出现拘束无语、拒访、不耐烦、怯场等情况,故与被访者建立融洽的关系很重要。

## (3) 与被访者建立合作关系

能否在短时间内与被访者建立起融洽的关系是访问成功的关键。从访问者角度看,一是要注意服饰的整洁、表情的自然大方;二是在谈话时尽量要主动、礼貌、友好;三是要尽量去发现被访者的优点,并诚恳地告诉他;四是要真诚地关怀被访人,以被访者可能关心的事为初始的话题。

## (4) 正式访谈

正式访谈的过程也是访问者不断提问的过程。在正式访谈中应注意:第一,访问者应始终保持中立的态度,不能以某种权威的身份随意向被访者发话;第二,访问要抓住中心及主题焦点;第三,要注意提问的顺序,不作大的跳跃式的提问,最好按时间先后来提问;第四,要根据访问对象的特点,灵活掌握问题提法与口气;第五,访问时要注意随被访者的表情来控制自己的表情与动作,不能毫无表情,也不能过于夸张;第六,要善于不动声色地尽量完整与准确地记录访谈的内容。

## (5) 结束访问

结束访问是完整的访问调查中不可缺少的一步。结束访问首先要求访问者能够得体地中断谈话,其次,访问员应向被调查者表示诚挚的谢意,这既是对被访者付出时间与劳动的一种尊重,也是为日后可能需重新调查或补充调查再麻烦被访者留后路。

### 3. 访问员的挑选与训练

用访问法搜集材料,访问员个人的品质、特征和能力对资料的真实、可靠与完整有很大的影响,因此,挑选与训练符合研究目标和任务的访问员,是任何访问法都不可回避的一项重要任务。

#### (1) 访问员的选择

访问员应具备的条件可分为两类,一类是任何研究的访问员都应具备的一般条件。主要有:第一,诚实的品质与一丝不苟的精神,这是访问员应有的最基本的品质;第二,对访问工作要有浓厚的兴趣和较强的适应工作的能力;第三,要有勤奋的工作态度与不怕遭受挫折的心理,如在遭受访问对象冷遇或拒访时应不灰心、泄气;第四,在访问调查的全过程中应有谦虚耐心、不卑不亢的态度;第五,要具备一定的表达能力。

另一类是由研究主题的性质、社区的类型及调查对象所规定的,也称为特殊条件,包括:第一,性别。对于访谈的重点对象,研究人员最好能亲自去,或请与访谈对象级别对等的有关领导支持参与访谈,效果就较理想。访问年长的男性,再好由男性访问员去,访问女同志时,则宜派女性为宜,以避免访问者与被访问者在性别上的差异而产生拘束。在访问生产、政治问题时,以男性访问员为主,而对家庭、婚姻等调查,女性访问员则更合适。第二,年龄。通常对有身份与影响力的领导或年龄较大的人进行访谈,派年龄大的访问员较合适,而对年轻人进行访问,派年轻访问员即可,但在较重大的政治与经济问题的研究中,年轻的访问员应尽量少用。第三,知识水平。一般而言,在访问比较复杂的问题时,对于访问员的访问技巧、访谈经验和其对被访者的反应程度就会有较高的要求,因此,在研究复杂问题时,宜挑选知识水平较高的访问员。

#### (2) 访问员的训练

经过训练的访问员与没有经过训练的访问员对研究结果的影响是不一样的。从资料搜集的角度讲,有经验的访问员当然是比较理想的。训练访问员的步骤包括:第一,由研究人员就研究的目的、



意义、整个调查的范围、访问的数量及每人的工作量、调查的步骤与每个阶段所需时间、报酬情况等介绍；第二，在指导访问员事先认真阅读调查员手册及其他有关该项研究材料的基础上，由研究人员对访问内容、回答类别、如何记录及访问应注意事项进行详细解答。第三，进行模拟访问，一般在访问员之间一对一互相访问，也可找实验点让访问员实际操作一遍，在模拟访问过程中，研究人员应从旁观察、协助与监督。第四，集体就模拟访问中的有关问题进行讨论。第五，建立监督管理访问质量的办法，如将访问员分成小组，按访问进程定期了解访问中出现的有关问题等。

### 三、问卷法

问卷法是社会调查中用来搜集资料的一种工具，它的形式是一份精心设计的问题表格，用于测量人们的态度、行为和其他社会特征。本世纪以来，问卷法常与抽样调查相结合，成为社会学研究的主要方式之一。

#### 1. 问卷的类型

根据使用方式不同分，问卷有两种主要的类型：自填式问卷与访问式问卷（或访问调查表）。自填式问卷一般采用邮寄或分发的方式送到被调查者手中，由被调查者在无访问员在场监督和帮助的情况下自行填写，然后通过邮局寄回或由调查员收回。访问式问卷由访问员根据问卷向被调查者提出问题，根据被调查者的回答进行填写。

#### 2. 问卷的结构

一份问卷的结构通常包括下述几个部分：

（1）封面信，即一份致被调查者的短信，因其常置于问卷封面或封口，故此得名。封面信中应说明调查者的身份，进行该项调查的目的与意义，被调查者的选取方式和他对填答问卷的作用，请求被调查者的合作及对回答保密处理的保证等。封面信质量的好坏会直接影响被调查者是否乐于接受调查或认真填写问卷，它是问

卷中一个非常重要的部分。封面信的语言应简洁、中肯,结尾要落款,还可附上调查者的通讯地址。

(2)引导语或填表说明。主要告诉调查者如何正确填写问卷,或提示访问员如何正确完成问卷调查工作的语句。填表说明应简明易懂,应假设回答者从未填过问卷。

(3)调查的问题,这是问卷的主体。

(4)其他资料。如问卷名称、编号等。

### 3. 问卷的设计

#### (1)问卷设计的步骤

问卷设计的步骤大体包括:第一,从研究的问题和理论假设确定需要测量的变量;第二,将变量操作化成若干具体的指标,围绕这些指标编制合适的问题;第三,确定问卷的形式和结构;第四,进行小规模试调查,以发现问卷设计中存在的问题;第五,根据试调查的结果进一步修改问卷。

#### (2)问卷的主题内容

主题内容通常包括:第一,关于回答者个人基本情况方面的,如年龄、性别、职业、文化程度、政治面貌、婚姻状况、收入、家庭规模等。第二,关于行为方面的,包括被调查者本人的行为及其了解的其他人的行为。第三,关于态度或看法方面的。第四,关于社会环境方面的。

#### (3)问题的形式

问卷中的问题主要有两种形式:开放式问题与封闭式问题。开放式问题不为回答者提供具体的答案,而由回答者自由回答。如“您为什么不与您的孩子住在一起?”。封闭式问题就是在提出问题的同时还给出若干可供选择的答案,要求回答者根据自己的情况进行选择。如“您认为要保证家庭成员对老人的尊敬,老人保留一些财产是不是很重要? ①重要②不重要”。

对封闭式问题答案的设计应注意穷尽性与互斥性,即一方面给出的答案要尽量包括各种回答的可能性,不能有所遗漏,另一方

面,各种答案也不能出现交叉重叠或包容的现象。如下面这个问题的答案就不是互斥的。

您的职业是什么(请在合适的答案号码上打√)

(1)工人 (2)农民 (3)干部 (4)教师

(5)医生 (6)司机 (7)其他

在所列的答案中,“工人”与“司机”就不是互斥的。

开放式问题的优点是可以允许回答者充分地发表自己的意见,获得的资料丰富、生动,且常能从中发现新东西。其缺点是答案五花八门,不易编码与统计汇总。封闭式问卷的优点是回答者回答问题较方便,且省时省力,资料统一标准,便于编码与统计。其缺点是资料失去了自发性与表现力,回答中的偏误难以发现。封闭式问题常用于大规模的调查,而在小规模探索性调查中,两者常一起使用。

#### (4)问卷语句的设计

问卷中的一切问题都是通过语言来表达的,因此,问卷语句的设计、构造会直接影响到问卷资料的质量。在进行问卷语句的设计时应注意:

第一,要避免诱导性问题。诱导性问题是以某种方式暗示回答者应该如何回答的问题。这类问题的用语不是中性的,一般带有某种倾向性。如用大多数人的看法或权威的话设计提问语句。例如:“大多数人认为利用工作职权之便给家庭成员各种好处是个严重的社会问题,您呢?”,又如“您认为中央号召的反腐倡廉运动很及时吗?”这些问题易使人产生心理上的顺向反应,即既然大多数人和权威都这样,我也不便持相反意见。

第二,敏感性问题的设计要注意技巧。所谓敏感性问题是那些人们不愿在公开场合谈论的问题。如性关系、避孕手段、私人财产、不轨行为、对国家方针政策的态度等,都属敏感性问题的设计应写明替回答者保密,同时,设计这类问卷时,还可以改变提问的形式,使回答者不以第一人称回答问题,而以第三人称回

答问题。如要了解育龄男女是否正确使用了避孕药膜,可以有二种提问:

a. 请问您是如何合作避孕药膜的?

b. 有人使用避孕药膜往往导致避孕失败,您是否可以猜测一下他们失败的原因是什么?

第二种提问方式就比第一种好。

第三,要避免双重问题,即避免在一个问题中询问二件事情。如“你的父母是干部吗?”这样的问题,使得父母中仅有一人是干部的无法回答。

第四,问卷语句中使用的概念要有明确的涵义,不能含糊不清。如“你对你们单位近年来的情况感觉如何?”是什么方面的情况呢,问题中没有给出来,这里应该明确提出是生产情况、福利待遇情况还是其他情况。

第五,应尽量避免否定式提问,因为被调查者往往会把句中的否定词漏掉,结果把否定句看成肯定句,造成答案错乱。如有人曾做过一次试验,向被调查者提出这样一个问题:“三 K 成员不可在公立学校中任教:a. 是 b. 否 c. 不知道”结果发现许多选择是的人其本意是认为三 K 成员可以在公立学校任教,而不是同意问题的陈述。

#### (5)问题的数量

一份问卷到底应包括多少个问题合适,这要视研究的内容、样本的特性、分析问题的方法及拥有的人力、物力等多种因素来决定,没有固定的标准。但一般来说,问卷问题不应太多。太长的问卷宜使被访者出现厌烦与畏难的情绪,影响填答问卷的质量与回收率。特别是邮寄问卷,一般应控制在回答者能在 30 分钟以内完成为宜。

#### (6)问题的次序

问题的次序对问卷资料的质量常常会产生一定的影响,故对问卷中的问题应进行精心的编排。问题次序编排的常用规则是:第

一,把简单易答及能够引起被调查者兴趣的问题放在前面。一般来说,第一个问题应是关于一般事实方面的问题,不需要什么考虑就可以回答。第二,容易引起回答者顾虑、紧张的问题应放在后面,如敏感性问题、个人收入方面的问题。第三,开放性问题应放在问卷的最后面。第四,行为方面的问题可先问,然后问有关态度或看法的问题,最后是有关个人基本情况的问题。第五,按一定的逻辑顺序排列问题。比如,时间顺序(从最早到最近或相反)、类别顺序等提问。同一类型的问题最好放在一起问。第六,应尽量使问卷中问题的形式、回答的方式等方面具有多样性,以吸引回答者的注意力。

#### 四、量表法

社会调查需要了解人们的态度、意见、观念、思想倾向和行为倾向等方面的内容。本世纪以来,社会科学家们发展出了用以测量这些内容的精密化的调查手段与方法。

##### 1. 量表的特点与作用

量表是适用于较精确地调查人们主观态度的测量工具,它由一组问题构成,用以间接测量人们对某一事物的态度或观念。所谓态度是指人们对社会事物的具有一定倾向性的、内在的主观评价。至于观念和信仰,它们是更持久、更稳定的态度,涉及到对人生、对世界的基本价值判断。在研究社会现象时,对人们态度和价值倾向的了解,可以说明和解释许多社会行为,并能对人们今后的行为作出预测。量表法就是对这些人们难以直接调查的态度或观念进行间接测量的一种工具。例如,我们想了解人们对个体户是否抱有偏见。在问卷中直接询问“你对个体经营者(即个体户)有何看法?(选择回答是:1. 很反感 2. 有些看不惯 3. 没好感也没恶感 4. 其他\_\_\_\_\_)”对这一问题,被调查者可以作出真实的回答,但结果不能精确地说明被调查者对个体户的偏见程度。而用量表测量则能得到更真实、更准确的信息。如以下 5 个问题组成的量表

(见表 14-1)就能了解人们对个体户的歧视态度。

表 14-1

提 问 项 目	是	否
(1)你认为,是否应当允许个体经营者加入共产党?	1	0
(2)是否应当允许个体经营者被选为人民代表?	1	0
(3)是否应当允许个体经营者担任乡、县级领导职务?	1	0
(4)你是愿意当个体经营者?	1	0
(5)你是否愿意你的子女与个体经营者结婚?	1	0

回答“是”,记为 1 分,回答“否”,记为 0 分。将这五个问题的回答分总加起来,就得出一个总分,它能较精确地度量人们对个体户的态度。

因此,量表是用多个问题来测量一个概念(或变量),是衡量某一概念(歧视程度、工作积极性等)的综合指标。一个量表犹如一把“尺子”,其作用在于精确地度量一个较抽象的或综合性较强的概念,特别是度量态度和观念的不同程度或不同倾向。量表可根据它测量的内容、用途及制作方式等而分成不同的类型,限于篇幅,下面仅介绍最为常用的一种:总加量表。

## 2. 总加量表

总加量表也称为李克特量表,它的最初形式是给出一组问题,请被调查者作出“同意”或“不同意”的回答。美国社会心理学家李克特对它作了改进,增加为“非常同意”、“同意”、“不一定”、“不同意”、“非常不同意”五种回答,各种回答分别记为 1-5 分,这样根据每个被调查者对各道题的回答分数的总加就得出一个总分,这一总分说明他的态度强弱或他在这一量表上的不同状态。

### (1)总加量表的形式。

下面我们以一个例子来说明总加量表的形式(见表 14-2)

表 14-2

## 人际关系量表

您是否同意下列说法,请在合适的回答栏中打√

提问项目	选择回答(只限选一项)				
	非常同意	同意	不一定	不同意	非常不同意
+1. 我在本厂有许多好朋友	5√	4	3	2	1
+2. 只要我需要,我相信本厂的大部分同事都会助我一臂之力的	5	4√	3	2	1
-3. 对周围的同事我很少关心	1	2	3	4√	5
-4. 我很难和本厂的人交朋友	1	2	3	4√	5
+5 我经常向本厂的人请教	5	4√	3	2	1
-6. 只有少数同事是开明的,多数人是具有偏见的	1	2	3	4√	5
+7. 大部分同事会为了集体的利益牺牲个人利益	5	4	3	2√	1
-8. 我相信大部分同事会背后中伤我,如果这样做可以使他们晋升的话	1	2	3	4√	5
-9. 我很少关心别人说什么,我只相信自己	1	2√	3	4	5
-10. 本厂大部分同事思想保守,他们怎么也不肯改变	1	2	3	4	5√
+11. 我跟所有的同事都是朋友	5	4	3	2√	1
-12. 我在本厂没有一个好朋友	1	2	3	4	5√

回答栏中的数字是每道题的每种回答的记分,它是根据测量的维度确定的。这一量表的维度是人际关系好为 5 分,坏为 1 分。根据这一维度来确定每道题的回答是否与维度的方向一致,如果一致,就记为“+”号,若“非常同意”这一道题,就表示“人际关系好”,可记为 5 分。如果与维度方向相反,就记为“-”号。它说明

“非常同意”，就表示“人际关系坏”，记为1分。12道题的总分是从12~60分，分越高说明人际关系越好。表14-2中打“√”是某一职工的回答，他的总分是 $5+4+4+4+4+4+2+4+2+5+2+5=45$ ，表明这一职工的“人际关系”较好。

## (2) 总加量表的制作

总加量表制作的程序是：

①根据所要测量的内容或变量(如“人际关系”)，收集大量与这一内容有关的问题，然后初步筛选出一组问题(一般为10~30个)作为初步量表。

②确定问题的类别和计分标准。回答的类别可分为五个等级，也可分为七个等级、三个等级或两个等级(“同意”与“不同意”)，计分的标准按测量的维度规定的方向和回答类别来制定。

③试调查，从调查对象中找一些人尝试回答初步量表，以便发现量表设计中的问题。

④计量各题目的分辨力，删除分辨力不高、即无法区分人们不同态度的题目，保留分辨力高的题目(一般为5~20个)组成正式量表。

利克特量表分辨力检查方式是：将试调查中得分最高的25%的人与得分最低的25%的人进行比较，然后计算出每道题的分辨力。例如，表14-2的“人际关系量表”是一个初步量表，在正式调查表，先对企业中的20分工人进行试调然后列出得分最高的25%人和得分最低的25%人，分别计算这两类人在每一道题上的平均分。例如总分高的5人在第1题的平均分为 $(4+5+5+4+5)/5=4.6$ ；总分低的5人在第一题的平均分为 $(2+2+1+1+1)/5=1.4$ ；两者相减 $(4.6-1.4=3.2)$ 即为第一题的分辨力系数。分辨力系数越小，说明这一题的分辨力越低，这种题目应删除。如第10题与第11题，其分辨力系数仅为-0.2与0.2，因此，除去这两题，保留其余的即组成了正式的“人际关系量表”。一个量表设计好以后，即可对调查对象作正式调查。



表 14-3

分辨力的计算

被调查者		题目												个人总分
		1	2	3	4	5	6	7	8	9	10	11	12	
总分高的 25% 人	工人 1	4	5	5	4	3	5	4	3	5	2	5	4	49
	工人 2	5	4	4	5	5	4	3	5	4	1	4	2	46
	工人 3	5	4	3	3	4	5	4	4	4	2	4	3	45
	工人 4	4	4	4	4	5	3	3	4	5	1	5	3	45
	工人 5	5	5	3	2	4	4	3	5	2	2	4	4	43
	工人 6	4	3	2	5	4	5	4	2	3	1	5	4	42
	.....													
	工人 15	2	4	2	3	2	2	3	4	3	1	4	4	34
总分低的 25% 人	工人 16	2	2	4	2	3	3	2	4	2	2	5	1	32
	工人 17	2	2	2	3	4	2	4	3	3	2	4	1	32
	工人 18	1	3	2	4	1	3	3	1	2	2	5	2	29
	工人 19	1	1	2	2	2	3	2	4	1	1	4	3	26
	工人 20	1	1	1	2	1	2	1	3	2	2	3	2	21
总分高的25% 人的平均分		4.6	4.4	3.8	3.6	4.2	4.2	3.4	4.2	4.0	1.6	4.4	3.2	
总分低的25% 人的平均分		1.4	1.8	2.2	2.6	2.2	2.6	2.4	3.0	2.0	1.8	4.2	1.8	
分辨力系数		3.2	2.6	1.6	1.0	2.0	1.6	1.0	1.2	2.0	-0.2	0.2	1.4	

### 第三节 社会统计资料的整理

广义地讲,资料整理即在资料分析之前所做的所有准备工作,既包括对定性资料的分类简化,也包括对定量资料的汇总和分组。考虑到现代搜集社会统计资料的特点,本文所谈的资料整理是狭

义的资料整理,主要指在对资料进行统计分析之前所做的准备工作,即如何将搜集到的资料转化成数字形式,变成计算机可以处理的数据文件。

## 一、资料的评价——信度和效度

社会统计资料搜集整理的目的是为了认识社会现象,了解社会现象的某些属性。确定某一特定分析单位的特定属性值或水平的过程,称之为测量。我们搜集到的统计资料,都是依靠某一种或某几种方法的测量而获得的。这里有两个问题,即:第一,某一测量手段(方法)是否具有前后一致性或是否可信,即信度问题;第二,某一测量手段(方法)是否成功地测量了想要测量的性质或是否有效,即效度问题。对搜集到的资料进行信度与效度的评价是资料搜集和整理过程中必不可少的一步。

### 1. 信度

信度即可靠性,系指测验结果的一致性或稳定性。如以某项工具重复测量某项持久性特质时,是否得到相同的结果。最简单的例子是体重的测量,如果二次测量体重都是60公斤,则使用的磅秤是可信的,若二次测量结果不一致,则使用的磅秤是不可信的。

对资料的信度评价基本上有两种方法:

#### (1) 交错法或折半法

这种方法等同于两个完全相同的磅秤来测量同一个人的体重。如研究人员设计两份问卷,每份使用不同的问题,但测量的是同一属性,让同一群被调查者回答。如果两份问卷测量结果是相同的,则测量就是可信的,否则就是不可信的。这样得到的信度,即为交错信度。

或者将上述两份问卷合成一份问卷(通常要求这两份问卷的问题数目相等),每一份作为一部分。然后考察这两部分的测量结果,看其是否相关。这种方法即为折半法。

#### (2) 重复检验法

即用同一种测验,对同一群被调查者进行前后两次测验,并考察两次结果间的相关性,即为重复检验法。

## 2. 效度

效度即正确性,指测量或其他测量工具是否成功地测量了所要测量的属性或程度。效度的定义含有两部分内容:(1)测量手段实际上正在测量所要测量的属性,而不是其他属性;(2)测量手段正确地测量了该属性。若某一测量手段满足了这两个条件,即为有效。由于效度评价的方法非常复杂,这里就不作详细介绍了。

## 3. 资料的整体评价

用某一种方法或几种方法搜集到的统计资料通常包括对分析单位各种属性的大量测量,其中一部分测量可能是可信的和有效的,其他测量则是不可信和无效的。要评价资料中每一项测量的信度和效度比较困难,通常的作法是只评价若干重要测量的效度。同样,在对资料的信度进行整体评价时,运用折半法相当困难,一般往往采用重复检验法。

## 二、资料整理的步骤

资料整理一般包括资料的审核、编码、建立编码簿、资料净化等几个部分。有些资料最终可整理成适合计算机分析的数据文件。

### 1. 资料的审核

审核是指对原始资料进行审查与核实的工作过程,目的在于保证资料的客观性、准确性和完整性。审核的原则有四条:(1)真实性原则,(2)标准性原则。要注意指标的定义是否前后一致,计量单位是否相同等;(3)准确性原则。要对资料进行逻辑检验,看是否前后矛盾;(4)完整性原则。要检查资料是否有残缺,残缺不全的资料会降低甚至失去研究的价值。

### 2. 编码

所谓编码就是给一个问题的每一类答案指定一个号码。这个号码可以是阿拉伯数字,也可以是英文字母等其他符号。如我们询

问被调者的性别,可以“1”代表男性,“2”代表女性。

编码的一般原则是:(1)明确地区分每一种答案类别;(2)任何一个可能的答案都必须有一个号码来代表。

编码一般分先编码和后编码两种。

### (1)先编码

在设计问卷的时候进行的编码叫作先编码。因调查尚未实施,先编码限于那些答案类别事先已知的问题。这些问题主要是封闭性问题,只可能有几种答案类别,或答案本身就已经是数字,不需要再作转换。例如,某一调查问卷中可能包含这样一个问题:

您的母亲是否出生在本市?(请在下列答案中选择一个)1. 是  
2. 否 3. 不知道

这是一个封闭性问题,可能的答案类别只有三个。对这个问题,可以进行先编码,在设计调查问卷时分别以数字1、2、3来代表“是”“否”、“不知道”这三个答案类别。

### (2)后编码

在调查已经实施、问题已得到回答之后进行的编码,叫做后编码。如我们询问被调查者的性别,如果采用后编码的方法,这个问题应该这样设计:

您的性别\_\_\_\_\_

### (3)无回答编码

对于没有回答的问题,我们也应该进行编码。无回答编码的一般原则是指定一个特别的号码,以保证明确地区分无回答与其他答案类别。如0或9或98或999等。

## 3. 建立编码簿

编码簿用以记录每一个数字所表示的实际意义,它相当于打电报的密码手册。调查者要根据它将问卷调查资料转换成能够统计、计算的数字,输入计算机,然后再根据它将计算机整理出来的结果转换成我们能阅读的文字。不妨举例说明之。如某部门为了了解老干部和老知识分子的状况,对其系统内60岁以上的11 892

人做过一次调查。下面是该调查表的一部分。

1. 姓名\_\_\_\_\_
2. 年龄\_\_\_\_\_
3. 业务专长\_\_\_\_\_
4. 家庭居住面积\_\_\_\_\_
  - a. 20 平方米以下
  - b. 20~45 平方米
  - c. 45~65 平方米
  - d. 65 平方米以上
5. 您是否退休?
  - a. 已退休
  - b. 延期退休
  - c. 暂缓退休
  - d. 该退未退

这个调查共提出了 60 个问题,收集到有效问卷 11 892 份。因手工整理工作量太大,故用计算机整理。下面是这次调查编码薄的一部分。

表 14-3 编码薄示例片段

格(列)	问题号码	项目名称	内容说明
1~5	1	姓名	答卷人姓名编号
6~7	2	年龄	被调查者年龄
8~10	3	业务专长	专业编码详见附表
11	4	家庭居住面积	1. 20 平方米以下 2. 20~45 平方米 3. 45~65 平方米 4. 65 平方米以上
12	5	是否退休	1. 已退休 2. 延期退休 3. 暂缓退休 4. 该退未退

从表 3-3 可以看出,编码薄包含 4 个主要项目。

(1) 每一项目的位数或格数。用计算机处理数据,其原始数据文件的每一行有 80 例(格),可以允许有 80 个字符。数据文件的输入有固定格式和自由格式两种。计算机系统是根据数据在数据文件中的位置来判定该数据是哪一个变量值。表中的 1~5 格,说明第 1~5 列是 5 位数,它的编码可以从 00001~99 999。我们的调查不超过两万人,故格够用了。留格的多少要考虑每一类别的子项数,不超过 9 个留 1 格,大于或等于 10 个,小于或等于 99 个留 2 格;依次类推。故第 4、5 题的回答都只有 1 个格。

(2) 问题号码、编码簿上的问题号码与问卷或调查表上的号码一致。

(3) 项目名称,有的也称变量名,指问卷或调查表中的问题或项目是什么。

(4) 编码的内容说明。指该问题或项目中各个数字的具体含义,有的教科书将此称为变量数值实际意义的说明。这里我们应注意的是关于开放性问题的编码。开放性问题要先解决分类类别问题。其步骤是将答案通阅一遍,然后概括、归纳,要注意穷尽性与互斥性原则。比如业务专长这个问题,通阅后归纳出了 29 个类别 250 多种专长。故给予其 3 个格,并用一附表专门说明这些类别的编码。如业务专长为文学评论,编码为 006。

#### 4. 资料净化

大多数情况下,搜集到的资料总会含有某些错误。在样本容量较小且询问问题较少的情况下,可以通过逐一阅读的方式修正。当样本容量大且询问的问题很多的情况下,可以采用资料净化的方法来检查资料的错误。资料净化一般有两种方式:一种是检查是否存在特定编码之外的其他数字。例如,被调查者的性别编码是数字 0 和 1,我们可以令计算机显示被调查者的性别构成状况,如果出现 0 和 1 之外的其他号码,就可以断定存在错误。另一种方法是运用变量之间的逻辑关系来检查资料的错误。如对 A 问题做了这样的回答,对 B 问题就有了限制。如果计算机显示结果对 B 问题的

回答没有限制,则可以断定存在错误。在实际资料搜集分析中,这两种方式常结合使用。

#### 5. 计算机在统计分析中的作用

在进行大规模的调查搜集资料的情况下,人工统计与汇总既费时,费力,又容易出错,而且对许多问题及变量间的关系无法计算。为此,专家们将计算机的分析与运用引入到了统计资料的分析中,并发展出了统计软件包。所谓软件包即一组一道工作的计算机程序。目前适合用来对社会统计资料进行统计分析的是统计软件包,有两种类型,一类用于一般的统计分析,叫做综合统计软件包;另一类用于特殊的统计分析,叫做特殊统计软件包。常见的综合统计软件包有“社会科学统计包(SPSS)”、“统计分析系统(SAS)”、“统计包(SP)”等。这些统计软件包可以在 IBMPC 微机上使用。综合统计软件包一般包含有多种统计分析功能。可以进行描述性分析,如频数分布、列联分析,也可进行相关分析,如回归分析、因素分析、路径分析、方差分析等。

附表1  
随机数字表

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	1	2	5	4	2	8	5	8	7	3	5	8	4	0	2	4	3	6	8	4	8	4	8	5	2	6	1	7	5	4	8	8
2	5	4	4	3	4	9	1	1	0	9	2	2	7	1	3	4	4	7	9	8	1	3	1	1	8	7	0	1	2	2	1	0
3	3	2	6	2	2	3	2	2	4	1	1	2	9	8	7	7	4	7	7	6	4	5	1	2	1	7	4	6	2	5	9	3
4	7	8	0	9	0	2	9	7	8	9	5	6	2	1	5	8	7	7	8	0	0	7	5	3	1	2	3	2	7	1	8	1
5	6	8	6	2	4	1	9	4	3	5	9	6	5	0	7	2	4	4	7	3	3	0	9	9	0	7	2	9	4	9	5	0
6	9	1	7	9	3	8	1	4	9	1	5	3	2	1	2	7	6	7	4	5	9	6	4	6	8	1	0	5	3	1	3	3
7	5	3	1	7	0	9	8	6	0	6	3	3	6	4	8	0	4	8	3	4	8	7	1	0	8	8	2	9	8	5	7	2
8	0	1	2	6	4	7	7	7	8	0	3	4	9	2	1	7	2	1	2	8	2	2	3	2	5	0	3	9	8	6	3	7
9	2	3	7	2	7	7	7	4	9	4	4	6	7	1	7	8	8	4	0	3	3	9	7	1	0	8	9	9	5	2	7	4
10	0	3	5	7	5	2	7	6	3	9	9	9	0	2	6	1	9	2	5	5	5	7	8	0	5	7	2	8	0	0	3	2
11	7	8	5	5	9	7	0	7	5	2	5	9	4	2	6	3	9	8	7	8	4	9	1	8	0	9	8	7	9	1	1	8
12	2	5	1	0	4	2	5	4	1	5	4	3	0	2	2	4	0	1	1	2	6	5	2	3	8	6	6	7	4	7	0	7
13	6	6	3	9	1	9	1	3	3	1	2	0	9	1	4	9	6	1	4	5	5	8	9	5	0	7	2	6	3	8	8	3
14	6	7	6	9	1	4	3	5	9	1	0	7	4	7	6	2	9	9	0	2	3	7	6	4	7	3	8	8	2	7	2	9
15	4	5	2	7	8	0	0	0	8	6	4	8	3	3	6	6	7	9	4	5	4	8	4	7	4	3	1	7	9	6	3	6
16	5	6	9	9	9	8	8	3	2	4	5	6	0	8	9	3	4	1	3	2	6	6	6	8	0	7	9	9	6	1	3	7
17	4	1	6	0	1	4	4	5	2	8	8	7	0	7	2	4	1	2	9	4	8	9	8	8	1	5	2	7	1	4	6	7
18	4	5	0	6	2	4	7	4	3	5	9	0	5	3	0	8	7	6	4	0	7	1	2	8	1	0	2	3	2	4	1	8



19	4 6 4 5 0	6 1 3 9 8	4 6 4 4 5	8 5 6 6 6	7 6 7 1 1	1 8 4 2 3	2 8
20	6 6 8 6 3	5 4 4 9 8	2 8 9 1 8	7 0 5 0 6	6 4 7 3 5	3 5 6 8 9	4 0
21	6 5 0 3 0	3 2 9 7 8	9 9 8 2 1	1 0 8 5 2	8 0 6 6 5	7 0 6 1 9	4 0
22	1 6 5 8 4	2 8 8 1 8	5 6 5 3 1	9 3 5 1 2	8 9 8 1 7	4 6 8 3 8	3 6
23	1 3 1 9 1	2 0 4 3 3	4 4 8 8 8	6 3 9 4 6	6 7 7 7 1	7 0 0 0 3	2 3
24	5 0 3 0 4	0 2 7 2 0	7 7 9 8 1	2 8 6 4 5	3 2 9 0 7	1 9 1 6 1	5 2
25	7 8 6 6 2	0 2 9 5 1	5 6 2 0 0	3 2 9 4 0	0 2 3 7 7	6 7 0 2 8	5 2
26	1 9 8 3 8	9 9 2 1 0	1 7 7 2 6	3 7 6 9 9	4 1 5 1 8	1 3 2 7 2	7 1
27	9 9 4 4 0	8 4 5 7 4	6 8 3 9 3	6 8 0 0 2	0 8 5 7 5	7 8 4 4 4	8 0
28	0 3 3 0 9	9 1 3 4 9	9 0 7 7 9	0 6 9 3 2	0 8 7 1 1	9 8 8 9 8	8 1
29	9 9 0 3 7	9 1 4 4 1	3 8 6 8 2	6 0 2 3 0	1 3 3 7 7	4 1 3 8 8	4 0
30	1 6 1 4 7	8 6 2 9 5	0 0 4 1 0	9 1 0 3 7	2 9 7 8 6	0 7 5 0 9	7 1
31	2 0 9 6 1	1 6 4 3 7	8 8 6 2 5	7 0 6 3 2	0 6 9 3 2	2 6 3 5 2	9 0
32	0 5 1 1 0	2 2 9 5 9	5 1 6 8 0	8 1 4 0 9	7 6 2 4 4	9 0 3 4 6	9 2
33	0 5 2 4 4	0 5 6 9 1	4 0 6 3 7	1 2 0 9 9	8 2 9 0 3	6 1 1 6 5	0 1
34	7 3 8 1 7	3 8 6 6 5	6 8 1 5 6	8 4 1 6 0	0 4 2 9 3	4 8 8 3 7	4 1
35	3 3 1 1 3	7 3 3 7 8	8 2 6 9 8	5 7 8 7 4	7 2 6 4 4	5 8 7 9 5	9 1

续表

36	6 8 7 4 2	5 3 4 7 4	8 5 9 5 9	6 9 0 8 6	2 7 0 1 4	9 6 7 1 5	8 8
37	8 9 8 7 3	1 2 1 3 6	2 8 0 3 7	2 1 0 5 9	6 3 3 9 8	9 7 3 4 2	1 8
38	9 2 0 5 2	3 5 8 0 3	9 3 3 1 9	6 1 6 1 2	8 3 9 7 4	3 9 0 0 1	8 7
39	9 7 4 9 7	9 8 3 7 0	7 6 9 7 9	1 1 5 3 0	8 1 2 7 4	4 7 4 1 8	9 5
40	2 1 8 3 2	1 0 9 2 8	7 4 5 7 3	3 1 5 6 7	7 7 6 4 4	9 3 9 9 9	1 9
41	6 9 2 6 3	0 8 5 2 0	7 9 3 3 3	0 4 4 3 2	9 5 2 4 0	3 2 7 9 6	4 0
42	3 3 7 3 3	5 6 7 0 3	7 1 5 9 3	2 3 9 2 3	7 2 5 0 8	5 7 8 5 8	6 9
43	9 7 7 1 5	5 4 2 4 7	1 5 5 5 2	7 3 7 6 3	3 1 6 7 3	6 7 9 4 3	9 9
44	9 3 5 3 5	5 7 6 5 4	7 4 0 1 9	0 7 2 7 4	6 9 9 3 3	9 2 0 7 2	7 2
45	5 7 6 1 6	3 0 1 3 5	5 8 6 2 0	5 3 0 1 2	6 1 9 5 8	4 6 1 2 0	4 6
46	6 8 5 0 8	1 2 2 4 4	5 5 2 9 4	0 9 9 4 5	9 6 8 8 3	5 8 8 8 3	1 1
47	6 6 1 6 6	7 6 0 4 9	3 8 0 0 6	6 0 3 9 1	8 8 9 8 4	7 5 3 7 4	0 2
48	2 6 3 3 4	2 5 5 8 7	5 5 8 4 3	4 4 3 3 4	8 9 9 2 1	2 6 0 0 5	4 7
49	6 8 3 7 4	8 9 8 9 5	2 7 3 5 2	6 6 5 3 6	7 7 0 3 9	9 4 1 8 4	5 4
50	4 5 2 3 7	7 7 2 3 8	8 8 7 2 4	3 1 3 3 0	7 1 1 5 4	8 9 7 6 6	1 8