

# 计算机软件与社会经济调查

李 兵

当前,以计算机为主要标志的新技术革命浪潮已深入到各个领域,社会经济调查也深受其影响,用计算机来处理数据资料已成为调查过程中不可缺少环节。计算机对社会经济调查技术的影响主要体现在如下几个方面:(1)承担了大量的数据处理工作,大大降低数据统计分析的工作量;(2)提高了数据资料的质量;(3)提供多种分析手段,为调查资料的利用开辟了广阔的天地;(4)促进了调查行为的规范化和标准化。

计算机技术应用于社会经济调查,主要承担了如下工作:(1)建立社会经济调查资料数据库,即将人工填写的调查表输入计算机,转变为计算机文件;(2)调查资料的整理与管理,如数据质量的检查,数据的增加、更新和检索等;(3)调查数据的统计与分析;(4)计算结果(如统计图表等)的输出;(5)调查数据的传递或通讯。以上计算机的各种功能都是通过软件来实现的,本文将从软件的角度出发来分析计算机在现代社会经济调查中的应用。

目前在社会经济调查数据处理中使用最广泛的软件主要有 4 大类:数据库系统、高级统计软件包、电子表格和字处理系统。

## 一、数据库管理系统 (Data Base Management System)

调查是观察、测量社会经济现象的手段,我们通过调查可以获得大量的原始信息,数据库系统的任务就在于对这些信息进行记录、整理和归类,将其规范化和数据化,并送到数据库文件中保存起来,对其进行管理。数据库系统中应用最广泛的便是 dBASE III 系统,以及与其兼容的后起之秀 FoxBASE 系统,目前在微机数据库管理领域中占统治地位的是由 Microsoft 公司在 FoxBASE 基础上推出的 ForPro。FoxBASE 系统的数据库是一种关系型数据库,即以二维表形式存贮和管理数据,一行信息称为一个记录(record);一个记录中各个独立的信息称为字段(field)。一般来说,数据库中的一条记录就对应于一份社会经济调查表格,一个字段就对应于调查表上的一个问题,在社会经济调查过程中,运用 FoxBASE 系统主要是解决如下几个问题。

### (一) 数据录入

数据录入就是将社会经济调查表上的信息输入计算机存贮起来,以便由计算机进行汇总、分析。任何调查,只要使用计算机来处理资料,数据录入过程就必不可少。数据录入是整个数据处理过程的第一步。

从录入方法来看,目前主要有两种:一是传统的手工键盘录入,即由录入员通过键盘将调查表上信息逐个键入计算机;二是使用现代的光学扫描技术,将调查表上的信息直接读入计算机。数据录入一向被视为数据处理的“瓶颈”,往往是数据处理过程中投入人力最多、使用设备最多和占用时间最长的一步。

一般来说,只要调查表的份数达到一定规模,无论是键盘录入还是光电扫描,录入过程中的差错都不可避免,数据录入质量的高低,直接影响到后续的数据整理、汇总、统计分析等环节的工作量和工作质量。因此,数据录入又是数据处理过程中最关键的一步。

提高键盘录入工作效率与质量的关键在于要编制一个好的录入程序,即为数据进入计算机提供一个好的“入口”。FoxBASE 系统在数据的输入方面为程序设计者提供了一套强有力的工具。录入程序的编制需掌握如下原则:第一,可靠性,即对输入数据的合法性进行检查,将人工可能出现的错误降到最低限度。第二,方便性,即向录入员提供良好的工作界面。要做到屏幕

格式清楚, 接近原始调查表格的卷面; 光标走向明确, 接近手工填写的顺序; 在清楚的前提下, 尽可能完整地显示表格, 以便录入员检查核对; 还要具备较好的人机对话功能。第三, 容错性, 即程序要具备自我维护功能, 要能应付录入员的各种误操作。在编程时必须有一个基本假设: 程序的使用者完全不懂计算机。第四, 灵活性, 程序要具备多种输入和查询的方式, 供用户选择。第五, 实用性, 要做到使用简单, 使从未接触过计算机的人员稍加培训即能上机操作, 为此甚至可以牺牲一些次要的功能, 尽量降低录入员的击键次数; 最好能使录入员实现“盲打”, 以提高录入速度。以上五个方面既互相联系, 又互相牵制, 往往不能面面俱到, 必须根据调查表的特征、录入员的素质、录入工作环境等因素, 确定编程的主要原则。

### (二) 调查数据的清理

在调查数据进入计算机, 作为一个数据文件保存之后, 一般都需要对数据资料进行审核、清理, 以保证数据的正确性。清理主要检查三个方面的问题: (1) 未回答。对于每一个应回答的问题, 表上都要有明确回答。(2) 回答不合要求。即对封闭性问题的回答超出编码范围; 对开口性问题的回答不合情理。(3) 回答前后矛盾。即对若干有逻辑关系的事实性问题的回答前后不一致, 以及对一些过滤性问题的回答不合要求。使用FoxBASE 进行数据清理, 可按照以上3个方面, 对有关字段列出逻辑关系式, 使用循环对调查表进行逐份检查, 然后对于发现错误的调查表进行修改。

### (三) 调查数据的管理

调查数据的管理, 即发挥FoxBASE 系统的分类、查找、检索、合并、增加、删除、编辑、显示、打印等多种操作功能, 来管理调查资料数据库。一般来说, 主要有两类工作: 第一, 以FoxBASE 系统为支持软件, 开发应用系统, 对一些需要经常查阅、更新、输出的调查资料进行管理。第二, 充分发挥FoxBASE 系统在数据库操作方面的长处, 从原始数据资料库中间接产生一些二级数据库, 为利用其它软件进一步处理社会经济调查资料作数据准备。

## 二、高级统计软件包 (Statistics Package)

任何软件都有其长处, 也都有其短处。FoxBASE 系统的长处在于数据库管理, 其短处则在于统计分析。从dBASE II到FoxPro, 系统的改进主要集中在数据库结构、函数、屏幕格式设计、变量处理、编程技巧等方面, 而其汇总、统计等功能则始终停留在初级阶段。FoxBASE 系统将数据的统计分析工作留给了其它高级统计软件包。

目前在国外使用最广泛的高级统计软件包主要有SAS、SPSS、SYSTAT 等。这些统计软件包的主要功能有: (1) 数据排序分类; (2) 图表和统计量的打印; (3) 简单的统计分析; (4) 方差分析; (5) 多变量的相关和回归; (6) 非参数检验; (7) 多变量分析 (如因素分析、判别分析、聚类分析等); (8) 数学模型拟合; (9) 报表输出。几乎包括了现代数理统计学中的所有统计分析手段, 为社会经济调查数据的统计分析提供了强大工具。

同FoxBASE 系统一样, 在统计软件包中, 供处理的数据是可以独立于软件包, 以文件的方式保存在磁盘或磁带上的, 在工作时按用户的指令将其读入 (或打开), 作为当前的活动文件 (active file), 然后对其进行处理, 系统将处理的结果以图表的形式向屏幕、打印机或文件输出。统计软件包活动文件的取得一般有4种方式: (1) 打开自身格式的数据文件; (2) 在系统内部直接输入数据; (3) 读入其它格式的数据文件; (4) 读入无格式的文本数据文件。在实际应用中, 很少采用在系统内部直接输入数据的方式, 数据输入不是统计软件包的长处。用得最多的是(3)、(4)两种方式, 统计软件包一般都具备很强的文件转换功能, 能够将其它格式的数据文件直接转变为系统活动文件, 从而弥补了自身的不足。

## 三、电子表格 (Electronic Spreadsheets)

电子表格的概念首次出现于Apple II微机上装备的VisiCalc 报表软件, 1983年Lotus 公司推出1—2—3后, 立即获得了成功, 目前Lotus 1—2—3与其后由Borland International 公司推出的QuattroPro、由Microsoft 公司推出的Microsoft Excel 并称报表软件的“三巨头”, 三大软件各具特色, 又相互兼容, 在激烈竞争中不断发展。电子表格的特点在于: (1) 兼备数据库、统计和绘图三类功能, 属于综合型软件; (2) 避免了较慢的DOS 调用, 采用全屏幕编辑的方式

直接将文本写到显示内存上,从而产生出快速应答的效果,对用户来说非常直观;(3)采用菜单工作方式,对于初学者来说,尤其容易使用。

电子表格的数据来源同统计软件包基本一致,也是4种方式,电子表格也具备很强的文件转换功能,能够读取多种其它格式的数据文件。但由于电子表格采用的是直接写屏的工作方式,所以用户也经常通过键盘直接输入数据,特别是对于一些规模不大的社会经济调查资料,用电子表格来建立数据库非常方便。

电子表格虽然集中了数据库、绘图和统计三大功能,但其数据库管理功能显然不及FoxBASE系统,统计分析功能则远不及统计软件包,唯在统计图的绘制方面电子表格有独到之处,为FoxBASE系统和统计软件包所不及。目前电子表格所具备的增强型图形和高级图形工具可使其图形输出质量达到出版物的水平,而且其图表以及文件连接功能也非常先进。相比之下,用统计软件包提供的绘图语句(如SPSS中的Plot)所绘制的图形就显得十分粗糙了。FoxBASE系统中FoxGraph虽然也可以绘制质量较高的图形,但由于FoxGraph使用方法十分繁琐,而且其图形的编辑功能不强,所以目前在实际工作中还很少用到。

电子表格的局限性在于:(1)虽然它能将数据库直接显示在屏幕上;但由于显示内存的限制,它不能处理规模较大的数据库。事实上,当电子表格的数据库的记录数约超过200个时,使用起来就已经不太方便了。(2)虽然它的数据输入采用了非常直观的直接写屏的填表方式,但这种方式的随意性太大,不利于数据输入和数据保存的规范化。(3)虽然它采用最简单、方便的菜单的工作方式,但由于它不能以程序方式工作,所以它不能自动处理复杂的问题。总之,电子表格所具备长处也正是它的短处所在。

#### 四、字处理系统 (Word Processing System)

字处理即在计算机上制作文档(document),目前国内比较流行的字处理系统有Word Star、WPS、WordPerfect和Word,一些应用软件(如FoxBASE、SPSS等)也自己配有字处理系统。字处理软件可以用来编辑文章、报告、信件等,也可以用来输入社会经济调查表上的各种信息,并将其作为一个数据块保存到磁盘上。

用字处理系统建立数据库,其基本要求是它所生成的这个数据文件必须是不带任何格式的纯文本文件,只有这样,数据块才能为其它系统读取或调用。我们所见到的字处理系统在功能和使用方式上各有差异,但每一种字处理系统都具备建立文本文件的功能。因此,原则上任何字处理系统都可以用来录入调查数据。然而,由于字处理系统不能对输入的信息作任何格式上的限制,也不能自动校验输入数据的合法性,所以我们一般不使用字处理系统来建立数据库。只是在一种情况下例外,当输入的信息是大块的文本时,采用字处理系统输入比在其它系统下输入要方便得多。如在FoxBASE系统下使用备注字段来存放开口性问题的回答、调查员后记等,用字处理系统输入显然要比在FoxBASE系统内部输入要方便。

对于社会经济调查来说,字处理系统的另一项功能便是可用来编写调查报告之类的文档,但这不属于本文要讨论的内容。

#### 五、数据文件的转换

充分实现数据共享是计算机数据处理的关键问题之一。目前绝大多数的计算机软件都具有以下一项或几项功能:(1)直接读写不同格式的文件;(2)提供把文件从一种格式转为一种格式的转换程序;(3)读写纯ASCII文件(文本文件)。以上4类软件都可以用来输入调查数据,但首选入口是FoxBASE系统,其次为电子表格;高级统计软件包一般不用来输入数据,但它具有很强的文件转换功能,能够读取或生成多种格式的文件;由于FoxBASE系统目前已成为较通用的数据库系统,所以它不主动与其它软件直接交换数据;字处理软件与其它软件的数据通讯一般都要间接通过文本文件,但有时也可以直接交换,如Microsoft Excel与Microsoft Word由于同属于Microsoft Office系统,因而可以直接交换数据。此外,在Windows环境下的软件都可以通过Windows提供的方式来交换数据。

(特约编辑:朱农)