

数据录入及数据管理的免费专业软件
由西藏卫生支持项目办向西藏所有卫生机构推荐使用

EpiData3.1

实用教程

扎西达娃 编著



A free software for data entry and management
Recommended by Tibet Health Sector
Support Program for all health agencies in
Tibet Autonomous Region



西南财经大学出版社
Southwestern University of Finance & Economics Press

数据录入及数据管理的免费专业软件
由西藏卫生支持项目办向西藏所有卫生机构推荐使用

EpiData3.1

实用教程

扎西达娃 编著



**A free software for data entry and management
Recommended by Tibet Health Sector
Support Program for all health agencies in
Tibet Autonomous Region**



西南财经大学出版社
Southwestern University of Finance & Economics Press

图书在版编目(CIP)数据

EpiData3.1 实用教程/扎西达娃编著. —成都:西南财经大学出版社,
2008.8

ISBN 978 - 7 - 81088 - 997 - 1

I. E… II. 扎… III. 数据管理系统—应用软件, EpiData3.1—教材
IV. TP317

中国版本图书馆 CIP 数据核字(2008)第 085742 号

EpiData3.1 实用教程

扎西达娃 编著

责任编辑:张 岚

封面设计:杨红鹰

责任印制:封俊川

出版发行:	西南财经大学出版社(四川省成都市光华村街 55 号)
网 址:	http://www.xcpress.net
电子邮件:	xcpress@mail.sc.cninfo.net
邮政编码:	610074
电 话:	028 - 87353785 87352368
印 刷:	四川森林印务有限责任公司
成品尺寸:	170mm × 240mm
印 张:	9
字 数:	145 千字
版 次:	2008 年 8 月第 1 版
印 次:	2008 年 8 月第 1 次印刷
印 数:	1—3000 册
书 号:	ISBN 978 - 7 - 81088 - 997 - 1
定 价:	18.00 元

1. 如有印刷、装订等差错,可向本社营销部调换。
2. 版权所有,翻印必究。
3. 本书封底无本社数码防伪标志,不得销售。

前 言

在科研实践中,通常我们先要做出一个详尽的调查表来采集科研资料。当这些资料积累到一定数量,我们会将它们输入计算机,使其数字化,方便保存和管理,最后再将这些数据进行统计学处理,找出其中规律性的东西。其中,使科研资料数字化是承上启下的一步。对于从事科研调查工作的人员来说,主要存在几个问题:①没有数据库的基础知识,不知道如何处理科研调查中纷繁复杂的数据;②在录入大量数据时容易出现错误;③英文基础不好,使用类似的没有汉化的软件(如 SPSS)显得额外费劲。而数据录入软件 Epidata3.1 具有简单、好用、实用的特点,是完全免费的软件,并已基本实现汉化,极其方便相关人员使用。

在西藏卫生支持项目办的大力支持下,本人得以重新编著了 Epidata3.1 软件的操作教程,在此向西藏卫生支持项目办的凯斯女士、来有文老师表示深深的感谢。在本书编写过程中,还得到了山东淄博的崔峰老师和西藏大学的达娃次仁老师的帮助,以及西藏金能电子科技有限公司的邓文、陈勇的技术支持,在此一并向他们表示感谢。

由于编者的水平有限、时间参促,书中难免有错漏之处,望广大读者和同仁给予指正。

编者

2008 年 4 月

目 录

第一章 简介	(1)
第一节 EpiData 发展简史	(1)
第二节 EpiData 3.1 的基础知识	(2)
第三节 “选项”设置	(5)
第二章 走马观花 EpiData 3.1	(10)
第一节 建立调查表文件	(11)
第二节 建立数据文件	(14)
第三节 数据导入和导出	(15)
第四节 建立 CHK 文件	(18)
第三章 建立调查表文件	(22)
第一节 字段编辑器	(22)
第二节 编码编辑器	(30)
第三节 定义字段名	(31)
第四章 数据文件的建立和修改	(36)
第一节 创建数据文件	(36)
第二节 根据修改的 QES 文件更新 REC 文件	(38)
第三节 重新定义字段名	(39)

第五章 建立 CHK 文件	(41)
第一节 添加/修改录入质控程序	(42)
第二节 数据核对窗口的设置	(45)
第六章 CHECK 的文件结构及命令列表	(52)
第一节 CHECK 的文件结构	(52)
第二节 CHECK 命令列表	(59)
第七章 运算符和函数	(90)
第一节 运算符	(90)
第二节 函数	(91)
第八章 录入数据	(99)
第一节 在字段间转换	(99)
第二节 在记录间转换	(100)
第三节 在关联数据库间转换	(100)
第四节 查找记录	(101)
第五节 查找字段和关联字段	(102)
第六节 滤过 (Filter)	(103)
第九章 数据库的管理和维护	(104)
第一节 数据处理	(104)
第二节 其他	(114)
第十章 错误的修复	(119)
第十一章 数据导入/导出	(124)
第一节 数据导出	(124)
第二节 导入数据	(128)
第三节 数据库的追加和合并	(131)
附录	(137)
参考文献	(140)

第一章 简介

第一节 EpiData 发展简史

EpiData 是一个免费的数据录入和数据管理软件。开发者是丹麦欧登塞 (Odense, Denmark) 的一个非盈利组织, 即 The EpiData Association。程序设计者为 Jens M. Lauritsen、Michael Bruus 和 Mark Myatt。

该软件目前有多种语言版本, 如丹麦语、挪威语、荷兰语、意大利语、中文、法语、西班牙语、俄语、斯洛文尼亚语、塞尔维亚语、波兰语、葡萄牙语、阿拉伯语、德语、罗马尼亚语、英语等。目前其中文版本为 3.1。

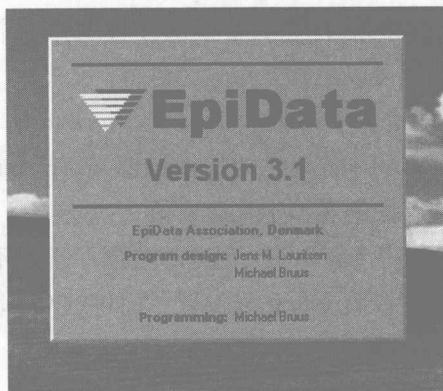


图 1-1

EpiData 的工作原理源自 DOS 版本的 Epi Info 6, 但工作界面为 Windows 版。EpiData 的安装、运行不会依赖系统文件夹中的任何文件, 也不会系统文件夹中安装或替代任何 DLL 文件。程序设置等参数被保存在名为 EpiDa-

ta.ini 的文件中。

理论上,该程序对录入的记录数没有限制。而实际应用中,记录数最好不要超过200 000~300 000(曾经用250 000测试过)。整个录入界面不能超过999行。对数值或字符串编码进行解释的文字长度最多为80个字符,编码长度最多为30个字符。

EpiData的研发工作最早由丹麦的Jens M. Lauritsen发起,最初是将其作为Funen县开展的“预防意外伤害行动”(Initiative for Accident Prevention)中的一部分。为什么会着手研发一种新的数据录入程序呢?

Epi Info 6具备所有我们需要的数据录入和管理功能。但是,随着视窗(Windows)程序的发展,很多使用者发现很难应付1990—1995年开发出的Epi Info的DOS模块。而商业化的程序一般不针对数据的管理,使用起来也没有那么简单,也不具备双录入核查的功能。

美国CDC的Epi Info工作组将Epi Info 6更新到Epi Info 2000。更新的Epi Info 2000采用了一种全新的工作策略,数据库不再采用过去简单的ASC II格式,而是启用Access数据库格式。

1999年末,Jens M. Lauritsen、Mark Myatt和Michael Bruus组成研发小组。Michael Bruus是一位专业的PASCAL(帕斯卡语言,一种高水平的计算机编程语言)程序设计员,编程工作主要由他完成。工作小组希望将EpiData开发成为一个简单、易用、独立的应用程序。这个程序不需要任何专门的数据库系统驱动(基于dll)。同时,他们希望从社会团体、个人以及其他捐助者那里得到资金支持,这样可以免费发布这个软件。

EpiData的操作原理是基于MSDOS版本的Epi Info 6。在开发EpiData时,基本的原则就是要保证EpiData创建的文件能够与Epi Info兼容,反之亦然。不过,两个程序在一些字段的类型上还是存在一些差异。基本上,EpiData和Epi Info v6非常近似。很多Epi Info 6创建的数据库基本不需要什么调整就可以用于EpiData,特别是当Epi Info 6只设置了简单的CHECK命令(例如允许值、重复、必须录入、跳转)时。

第二节 EpiData 3.1 的基础知识

EpiData主要用于数据输入。它可以将我们在科研调查之中所使用的观察表格“计算机化”,使计算机上的表格可以与我们的观察表完全一样,这样数

据输入就变得直观、简便。

一、EpiData 3.1 的下载与安装

<http://www.EpiData.dk> 是 The EpiData Association 的官方网站。我们可以进入该网站，并打开 Download (get files) 查找 Chinese (simplified)，点击“CN SETUP”下载。

下载完毕，直接双击所下载的“setup __ EpiData __ cn. exe”文件安装即可，也可以直接拷贝 EpiData. exe 文件到计算机中，同样可以运行。

二、EpiData 3.1 的组成

EpiData 由三种基本文件组成：

- (1) “. qes” 文件(调查表文件)。它的作用是定义调查表(问卷)的结构。
- (2) “. rec” 文件(数据文件)。它包括数据以及已经定义好的编码。
- (3) “. chk” 文件(数据录入核查文件)。它是数据输入字段的有效性规则。

字段是 EpiData 3.1 中最重要的概念，每个字段就是我们科研调查表中的一个小项目。

字段包括：

- (1) 字段的名称。它是给计算机程序识别的字段的名字，如 V1、V2 等，当然如果我们不去专门定义的话，EpiData 3.1 可以自动给字段定义一个名称。
- (2) 字段的文本描述。它负责解释这个字段在我们科研调查表之中所代表的具体观察指标，如年龄、性别、职业、婚姻状况等。

(3) 字段的数据变量。根据字段的性质，可以规定该字段的变量及变量长度。EpiData 3.1 支持的数据字段有：数字、文本、大写文本、欧洲日期、美国日期、布尔函数、自动 ID 号、声音提示字段等，可以基本满足科研调查的需要。

三、EpiData 3.1 的运行界面

运行 EpiData. exe 会出现程序运行界面(图 1-2)。

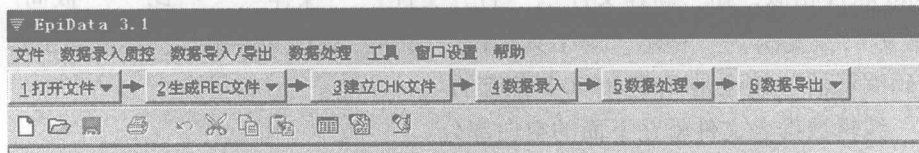


图 1-2

第一行是 EpiData 3.1 菜单栏, 所有命令都可以在其中找到:

(1) 文件。下拉菜单的项目有: “生成调查表文件 (QES 文件)”、“打开 EpiData 文件”、“选项”、“近期使用的文件名”及“退出”等。

(2) 数据录入质控。下拉菜单的项目有: “增加/修改录入质控程序”。

(3) 数据导入/导出。下拉菜单的项目有: “数据录入/编辑”、“根据 QES 生成 REC 文件”、“备份 REC 文件”、“导入数据”、“数据导出”、“纵向追加记录/横向合并字段”。其中“导入数据”下列有“从 TXT、DBF、SAS 文件导入”; “数据导出”下列有: “导出为 TXT 文件 (文本文件)、DBF (dBase 文件)、XLS (Excel 文件)、DTA (Stata 文件)、SPS (SPSS 文件)、SAS (SAS 文件)”, “导出为另一个 REC 文件”。

(4) 数据处理。下拉菜单的项目有: “显示 REC 文件结构”、“关于数据录入的备忘录”、“数据一览表 (每条记录为一行)”、“数据一览表 (每条记录为一段)”、“数据简单分析表”、“一致性检验 (对调查表双录入后的差异比对)”、“可靠性检验 (对录入后的 REC 文件进行质控)”、“按字段统计记录条数”。

(5) 工具。下拉菜单的项目有: “根据 REC 文件生成 QES 文件”、“清理 REC 文件”、“重建索引”、“根据修改的 QES 更新 REC 文件”、“字段重命名”、“编辑 REC 文件标记”、“复制 REC 文件结构”、“颜色编码表”、“准备双录入实时校验”、“重编码 REC 文件”、“清除所有录入质控程序”、“创建压缩文件”、“释放压缩文件”等。

(6) 窗口设置。下拉菜单的项目有: “水平平铺”、“垂直平铺”、“重叠窗口”、“显示所有窗口”、“工具栏”等。

(7) 帮助。下拉菜单的项目有: “内容”、“字段类型”、“EpiData 介绍”、“EpiData 主页”、“EpiData email 信息”、“关于 EpiData”等。

第二行是 EpiData 工作过程工具栏, 有 6 个项目: “建立调查表文件”、“生成数据文件”、“添加核对命令”、“数据录入”、“其他功能”、“数据输出”。

第三行是 EpiData 常用的几个快捷工具按钮, 大多数是与我们常用的 Word 软件相似, 如“新建文件”、“打开文件”、“保存”、“打印”、“返回”、“剪切”、“复制”、“粘贴”等工具按钮, 其中“REC 文件预览、字段编辑器、编码编辑器”按钮是 EpiData 3.1 独有的。

编辑数据表文件就在下面的空白部分。

第三节 “选项” 设置

作为基础知识的一部分重要内容，我们有必要先行讲解“选项”的设置。

当所有窗口关闭或编辑器处于激活状态，我们可以从菜单“文件”中选择“选项”进行相关的设置。所有的设置将保存在 EpiData. ini 文件中，因此，即使程序关闭，再次启动时，设置依然相同。

一、QES 文件显示

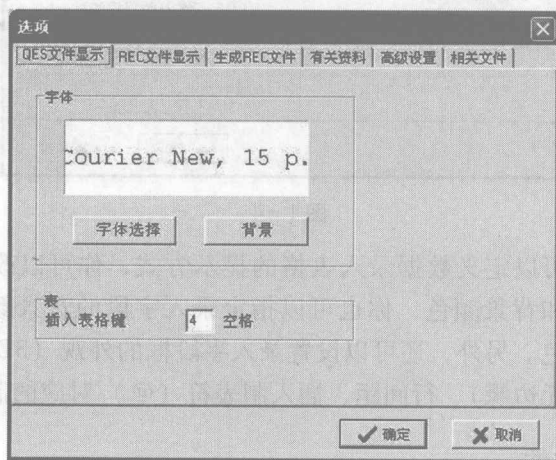


图 1-3

在这里，你可以修改编辑器窗口中的字体和背景颜色。设置完毕后，点击“确定”关闭选项设置窗口，所有当前开启的编辑器窗口都会更新为新的字体和背景颜色。

按 Tab 键时，你可以插入的空格数。EpiData 3.1 中不能使用“Tab”字符，所以当你打开一个文件，或从另外一个程序中拷贝文本，粘贴到 EpiData 的编辑器中时，程序会自动用空格替代原有的“Tab”字符。

另外，如果你发现用你的 EpiData 3.1 编辑器打开的含有中文的调查表文件显示为乱码，你可以尝试在这里更换一下显示的字体，很有可能是因为你的系统中不具备指定的字体，才导致乱码的出现。

二、REC 文件显示

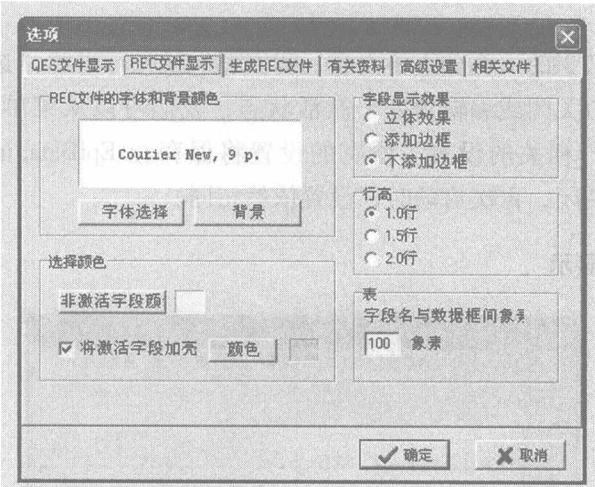


图 1-4

在这里，你可以定义数据录入表格的显示方式，你可以改变数据录入表格中显示的字体和背景颜色，你也可以指定录入字段的背景颜色，以及当字段被激活时的颜色。另外，还可以设置录入字段框的外观（3D 形式、平面式有边缘、平面式无边缘）、行间距、插入制表符（@）对应的制表位置间距。

三、生成 REC 文件

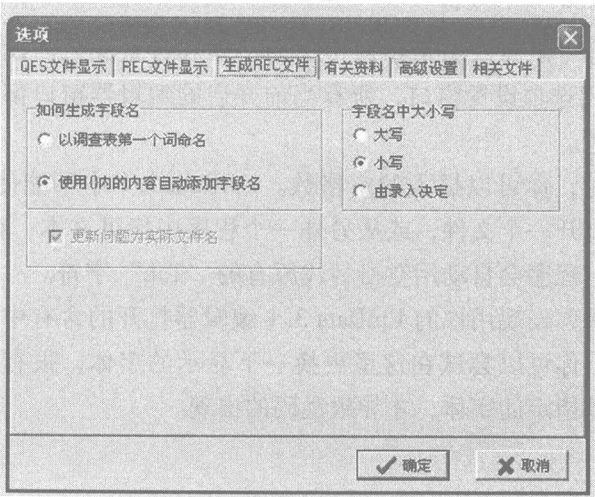


图 1-5

在这里，你可以设置字段的命名方式，以及字段名是大写、小写还是按 QES 文件中键入的实际情况显示。

四、有关资料

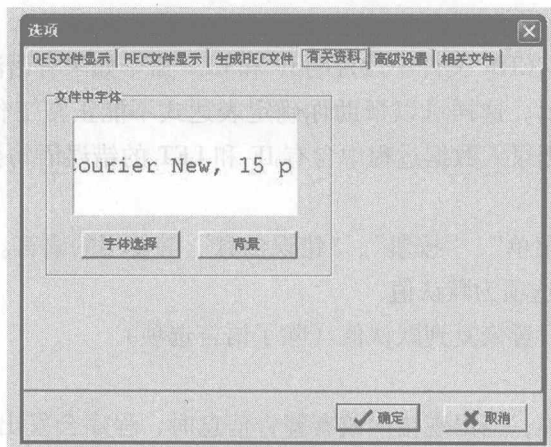


图 1-6

在这里，你可以定义输出报表（例如 File Structure、List Data、Codebook 等）的编辑器窗口的外观（字体、背景）。

五、高级设置

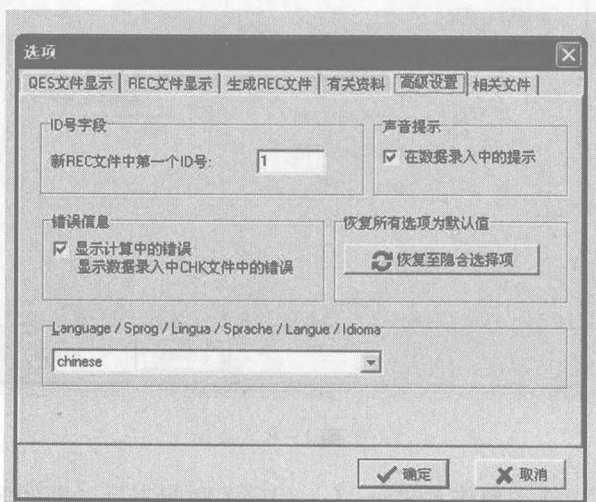


图 1-7

高级设置的内容有：

1. ID 号字段

如果新数据库中含有 IDNUM 字段，则该项功能可以定义第一个号从几开始。

2. 错误信息

勾选此项后，CHK 文件中设置的 IF 和 LET 命令如果有语法错误，程序会弹出错误信息提示。这样可以帮助你确定表达式不能正常工作的原因。如果没有勾选此项，则录入数据过程中含有 IF 和 LET 的错误语句会被忽略。

3. Language

选择显示“菜单”、“按钮”、“错误信息”等使用的语言。

4. 恢复所有选项为默认值

将所有选项设置恢复到默认值（除了语言选项）。

5. 声音提示

如果勾选此项，当出现错误或有警告信息时，程序会发出标准的嘟嘟声。可参见 CHK 文件中关于 BEEP 命令的讲解。

六、相关文件

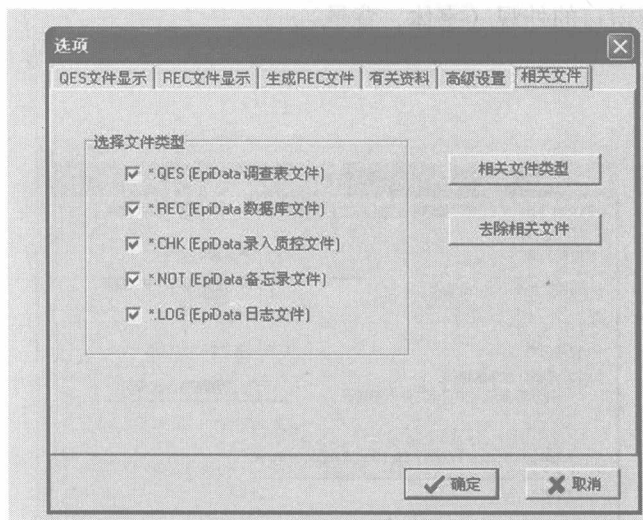


图 1-8

将 EpiData 文件类型与 EpiData 3.1 程序关联起来后，即使 EpiData 程序尚未启动，当你在资源管理器中直接双击一个数据库文件（.rec）时，也可以

激活 EpiData 程序，打开数据库。关联的文件类型有专门的图标，你可以很容易地辨别出可用 EpiData 程序打开的文件。

表 1-1

图标	文件类型
	调查表文件 (QES 文件)
	数据文件 (REC 文件)
	数据录入核查文件 (CHK 文件)
	备忘录文件 (NOT 文件)
	LOG 文件
...	...

第二章 走马观花 EpiData 3.1

EpiData3.1 是简单易学的统计软件。在本章,我们先以某单位针对“大骨节病”所设计的调查表(图2-1)为例,从“建立调查表文件”、“建立数据文件”、“数据导入导出”到“优化数据表”,将 EpiData3.1 软件的使用过程较粗略地学一遍,使初学者对 EpiData3.1 软件的使用有个整体概念。在以

“大骨节病”家系调查表

____地区____县____乡____村

今日日期_____

村编码:

家系编码:

个体编码:

A. 一般情况

1. 姓名_____

家庭总人数_____人

是否为户主_____ (若“是”,请跳答第3题)

2. 与户主的亲缘关系_____

3. 性别____ “Y”男 “N”女

4. 民族_____

(1)藏 (2)汉 (3)回 (4)蒙

5. 出生年月_____(年月日) 周岁____

6. 婚姻状况_____

(1)未婚(2)已婚(3)丧偶(4)离异

7. 职业_____

(1)农民(2)牧民(3)工人(4)教师(5)家庭主妇(6)学生

(7)医生(8)木匠(9)画匠(10)行政干部(11)其他_____

.....

9. 居住本村年限_____

图2-1 一张原始的调查表

后的章节，再对 EpiData3.1 软件使用的各个过程进行详细的讲解。EpiData 的使用可粗略地分为四个步骤：建立调查表文件、建立数据文件、数据导入及导出、建立 CHK 文件。

第一节 建立调查表文件

点击“打开文件”按钮，选择“建立新调查表文件”，命名为“大骨节病家系调查表3”并保存。这时生成的是“.qes”文件。

根据原始表格在 EpiData 输入相关内容，将文字部分全部录入。在“地区”之前，点击“选择字段类型”按钮，在选项中选择“文本”，设置为文本，定义长度为8个字节（根据实际情况，定义字节长度）。注意：一个汉字的长度需要2个字节。点击“插入”按钮（图2-2）

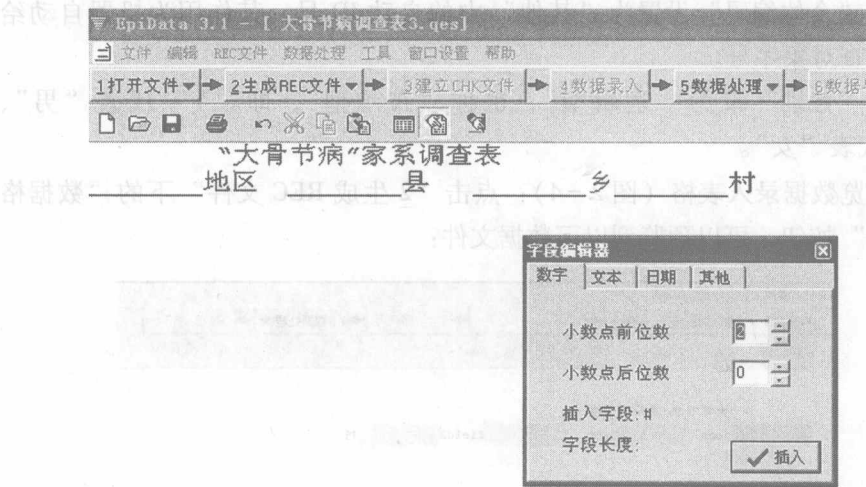


图 2-2

依此类推，将其他调查指标输入并保存（图2-3）。值得说明的是：“今日日期”为“字段编辑器”中“日期”的“自动插入日期”，它的作用是将将来我们在录入数据时使计算机可以自动给出当日日期（计算机上的日期，因此机器上的日期也应设置正确）。其格式为“年月日”，不用我们手工录入。将“村编码”设置为“数字”，长度视实际情况而定。