

Epidata 3.1 使用教程之快速入门篇(90 分钟速成)

第 1 节 建立调查表文件

我们为什么使用 [Epidata](#)?

因为 Epidata 是“免费的数据录入和数据管理软件”，另外最主要的，我想是因为它“简单易学、数据录入功能实用”等特点。关于这个软件的详查介绍，就不多说了，网上有很多。

现在的 Epidata 中文版本是 3.1，[下载地址](#)。

OK，简单介绍一下，废话少说进入正题。

一、建立调查表文件

使用 Epidata 录入数据的第一步，是先建立数据库。这时我们需要先写一个后缀是.qes 的“调查表文件”，如下图：

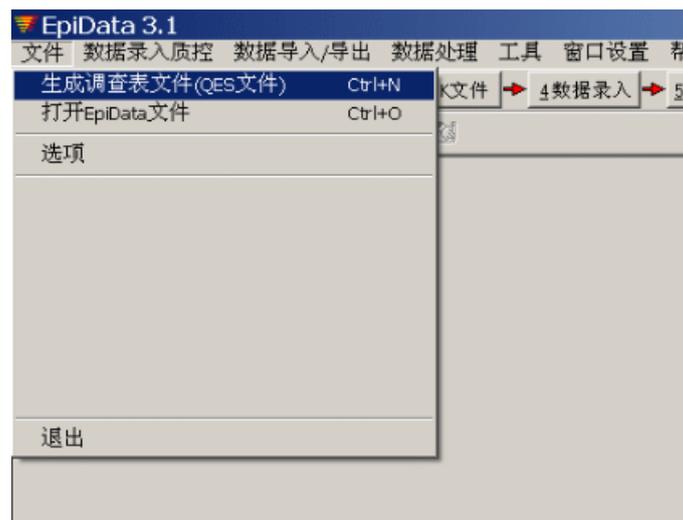


图 1-1 建立调查表文件

这个调查表怎么写呢？看下面的这个例子：

- 1、药物编号({DrugNum}):@####
- 2、患者姓名缩写({Pname}):@_____
- 3、就诊日期({date}):@<yyyy/mm/dd>
- 4、门诊 ({outp}) <Y> 住院({inh}) <Y> 住院病案号({pn}):_____
- 5、性别({sex}):# ①男 ②女

说明：

第 1 句中

“药物编号”只是起到提示作用, 在由.qes 文件生成数据库文件时不编译, 就是说在数据库里面它还是显示“药物编号”;

后面括号{}里的“DrugNum”, 是字段名, 当然这个字段(或者说变量)记录的就是药物编号了, 用{}定义字段名是 Epidata 字段命名的一种方式(还有另一种, 这里就不说了), 当然如果不加的话也可以, Epidata 会根据系统的设置自动生成字段名(象这种: n1aa,n2,n3ae 等等);

再后面的@###: @表示一种对齐方式, 先不用管它, 后面会讲到;

###是用来定义前面那个 DrugNum 字段的属性的, #代表一位数字, ###就代表 3 位数字, 也就是说药物编号最大也就能输入 999 了, 不能再大了;

总结第 1 句的含义: 定义了一个字段“药物编号”, 字段名是 DrugNum, 用于记录 3 位数字形式的数据。(刚开始有些费劲, 呵呵, 以后就越来越轻松了!



开始有些费劲, 呵呵, 以后就越来越轻松了!)

第 2 句:

与第 1 句不同的是, 患者姓名缩写肯定是字母了, 这里用下横杠“_”来定义。下横杠“_”是用来定义字符串的, 一个字母用一个“_”表示(如果是汉字, 要占两个“_”)。**这一句的含义就是: 定义了一个字段“患者姓名缩写”, 字段名是 Pname, 用于记录字符形式的数据。** (以后就轻松了: -)

第 3 句:

<yyyy/mm/dd>是定义日期格式的字段的, 写法是固定的, 很简单吧? 这一句将“就诊日期”的字段名设为“date”, 字段属性设置为 4 位年 2 位月和 2 位日形式的日期数据。

第 4 句:

<Y>代表另外一种格式的数据: 布尔逻辑变量, 它的值只能是 Y 或 N (在录入时也可输入 1 或 0, 系统自动将其变成 Y 或 N); 对于一名患者, 一般情况下只有门诊病人或住院病人两种情况, 非此即彼, 所以设置为布尔逻辑变量。**这一句的含义: 定义门诊与住院两种情况, 如果是住院病人, 还要填写住院病案号。**

第 5 句:

这句里面没有陌生的语句了, 为什么还要写出来做例子呢。因为这是一种常用的做法。比方说我们要记录一个患者的病情严重程度, 按标准分为四级, 0 无症状 1 轻 2 重 3 极其严重, 那么就可写成第 5 句这种形式:

严重程度({degree}):# 0 无症状 1 轻 2 重 3 极其严重

这里面只有#是在数据库中需要编译的, 其它都将显示为提示信息。**这里实际上是用一位数字存储等级资料的信息。**如果分级多, 可以设成两位数字(即两个##)。

当然, 对于第 5 句中的“性别”, 也可用第 4 句中那个布尔逻辑变量, 写成这样:

性别({sex}):<Y> 男

或者:

性别({sex}):<Y> 女

因为一般情况下性别也是非男即女（当然也不排除有人妖的可能，，所以也可写成：1 男 2 女 3 其他，用#来定义）。

事实上，上述例子，基本上把 EpiData 中的所有常用的变量类型都用上了。

作为练习，把上面那 5 句拷贝到 EpiData 的编辑区，保存为 .qes 后缀的调查表文件，即建库文件，。

第 2 节 生成数据库文件

二、生成数据库文件(.rec)文件

把第 1 节中那 5 句 COPY 到 epiData 中，如下图：

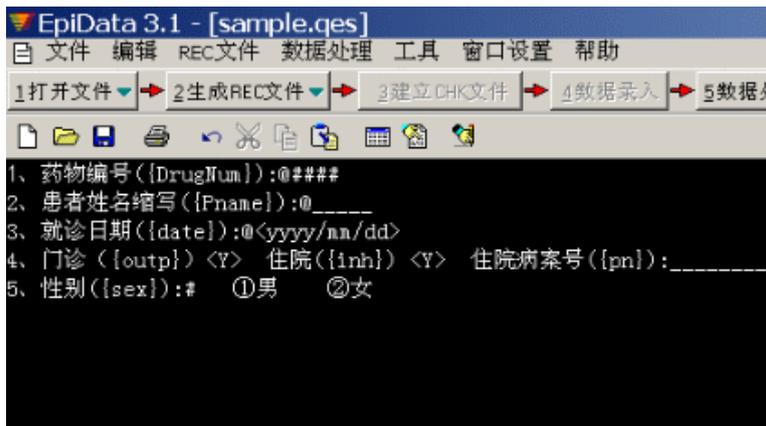


图 2-1 QES 文件编辑窗口

然后点击“生成 REC 文件”，如下图：

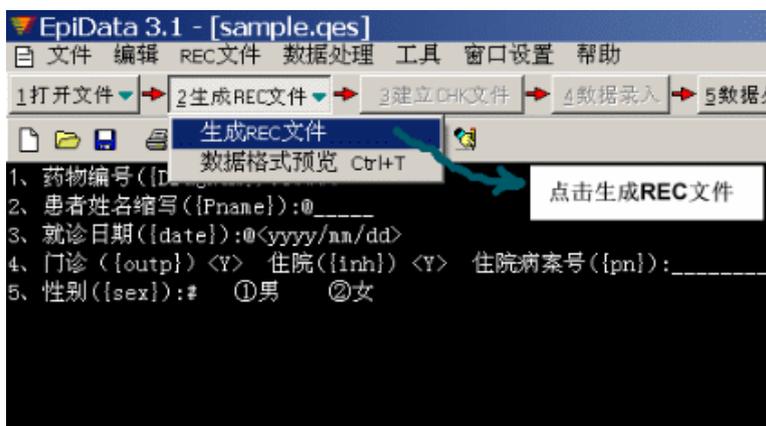


图 2-2 由 QES 文件生成 REC 文件

一般不用管它，你的 .qes 文件在什么位置，就把 .rec 放到什么文件夹中，所以下面基本上都是一路“确定”下去了：



图 2-3 生成 REC 文件

最后出来的这个文件标记，一般是不用设的（具体它有什么作用我还真没仔细研究过），最后一个“确定”点了以后，你的数据库文件（以.rec 格式保存）就生成了。

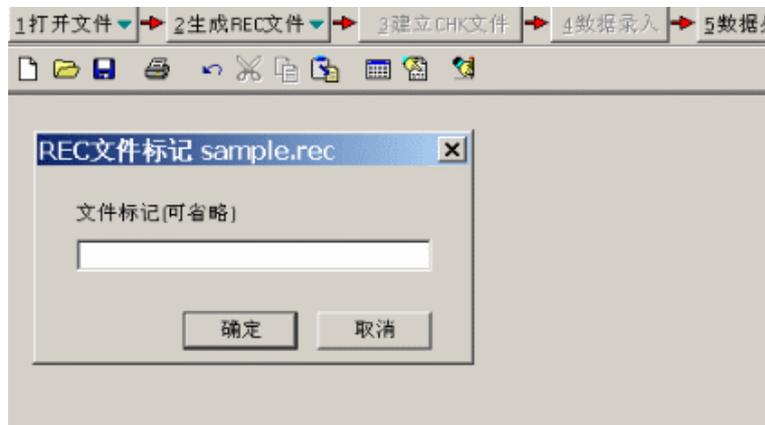


图 2-4 生成 REC 文件过程中的文件标记

好啦，来看看我们的成果吧！

点击 EpiData 工具栏中的“数据录入”菜单，选择“数据录入”菜单，如下图：

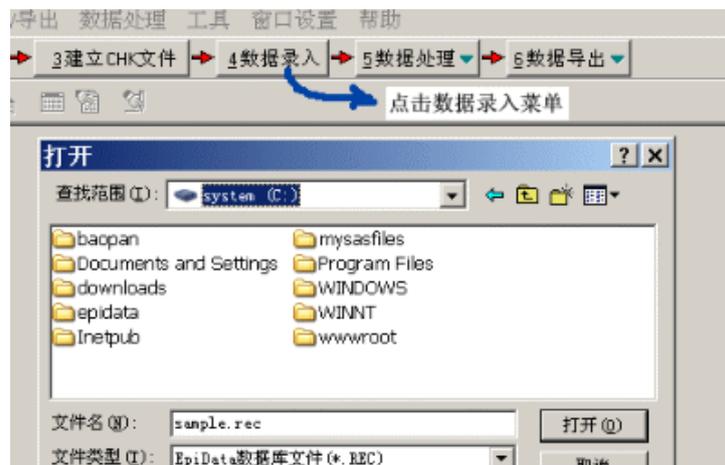


图 2-5 打开数据库文件

找到先前生成的那个 sample.rec 文件，然后打开，呈现在我们面前的就是数据录入的界面了：



图 2-6 数据录入界面

第 3 节 数据录入的控制(1)

前两节我们了解了如何编写调查表文件，如何用这个文件生成数据库及数据录入界面的显示。

这一节，我们看看数据录入的控制。

问题提出：**为什么需要“控制”？**

还是先前的那个例子，

- 1、药物编号({DrugNum}):@####
- 2、患者姓名缩写({Pname}):@_____
- 3、就诊日期({date}):@<yyyy/mm/dd>
- 4、门诊 ({out}) <Y> 住院({inh}) <Y> 住院病案号({pn}):_____
- 5、性别({sex}):# ①男 ②女

其中编号为 4 的那一行，有“门诊”“住院”和“住院病案号”三项，这三项中最多只有两项需要我们来录入数据，因为门诊和住院这两种情况是相斥的，一个病人如果是门诊病人，那就不是住院病人，病案号当然也没有了。

所以，为了提高数据录入的效率，我们需要对录入过程进行控制：对于上面的这种情况，我们的预期控制效果是：

在录完就诊日期后，光标到“门诊”项，这时，如果这个病人是门诊病人，那么我们录入 1 或者 Y（因为这个变量是布尔型，其值只有 1 或 Y，0 或 N，键入其它的值都是无效的）之后，光标不是到“住院”项，而是跳转到“性别”，因为门诊病人没有病案号。

当然，对数据录入的控制还远不只这些，还包括字段允许值的设定、是否必须录入、是否允许重复数值等。

为了实现这些“控制”过程，我们需要建一个与.rec 文件同名但后缀为.chk 的文件。在启动 epidata 后，在工具栏上就能看到**建立 CHK 文件**的按钮。

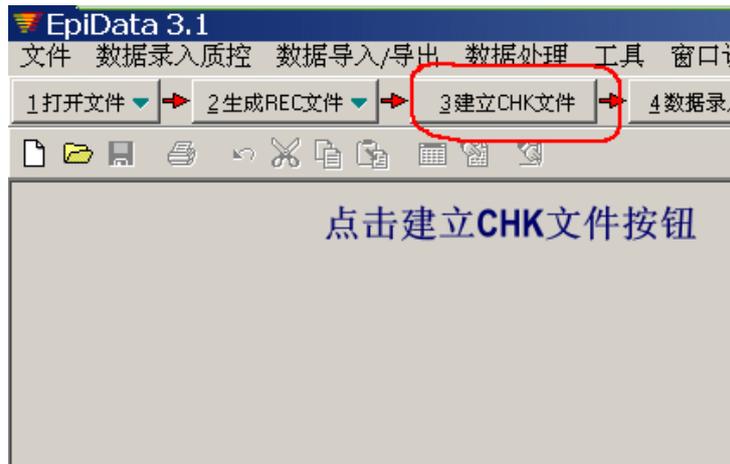


图 3-1 建立 CHK 文件

选择我们先前生成的 sample.rec 文件：

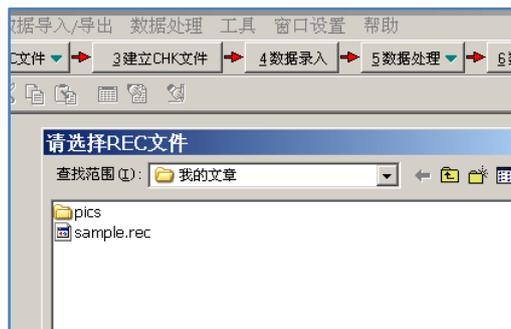


图 3-2 选择要建立 CHK 文件的数据库文件

点击打开后，出现编写 CHK 文件的页面：

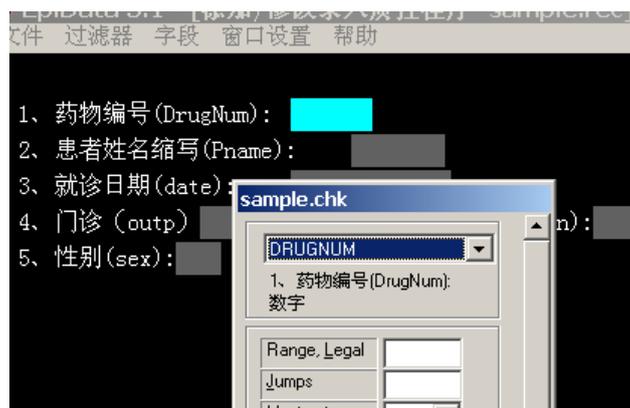


图 3-3 CHK 文件编辑器

简单的解释一下：

Epidata 的 CHK 文件，命令有很多，这个编辑窗口，仅列出了比较常用的几个，其它更为复杂的功能，需要点击“编辑”按钮，以文本编辑方式打开 CHK 文件来编写 CHK 命令。这些命令的使用，大家可以参考《[Epidata 3.0 使用手册](#)》。

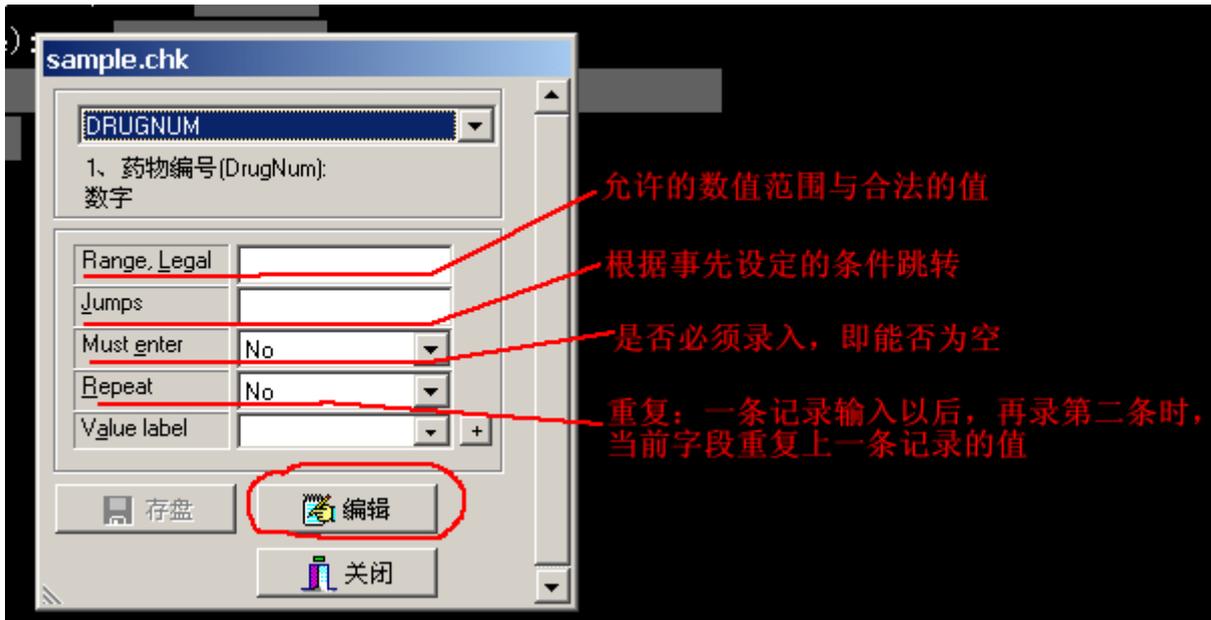


图 3-4 CHK 文件编辑器的选项卡

上图中对 CHK 选项卡中的 5 个项目作了简单的说明，下面逐一进行介绍：

1、字段的允许数值范围与允许数值

对应的 CHK 命令为：**Range 和 Legal**

“控制”作用：在录入时，如果录入的数值不在这两个命令定义的允许数值（范围）内，系统将报错，需重新录入“合格”的数值。

这是 CHK 文件中常用的两个命令，其作用稍有不同，Range 限定的是字段数值的“允许范围”，比如我们可以键入“1-8”，代表该字段的数值是从 1 到 8 且连续的 8 个数值；而 Legal 则限定字段的允许数值，对上面这种情况，用 Legal 命令的话，我们需要键入“1,2,3,4,5,6,7,8”，如果实际情况是既有连续的数值，又有不连续的，那这两个命令可以一起用，比如键入“1-5,7,8”，表示这个字段共允许 7 个数值，包括 1 到 5 连续 5 个数，还有“7”和“8”这两个数。如果录入时输入了 6，系统会报错，提示“非法录入”，同时还提示应该输入的允许数值（范围）。如下图：



图 3-5 (Range 和 Legal 这两个命令的作用就是这个)

还记得上一个图中那个“编辑”按钮吗？我们点击它看看里面是什么：

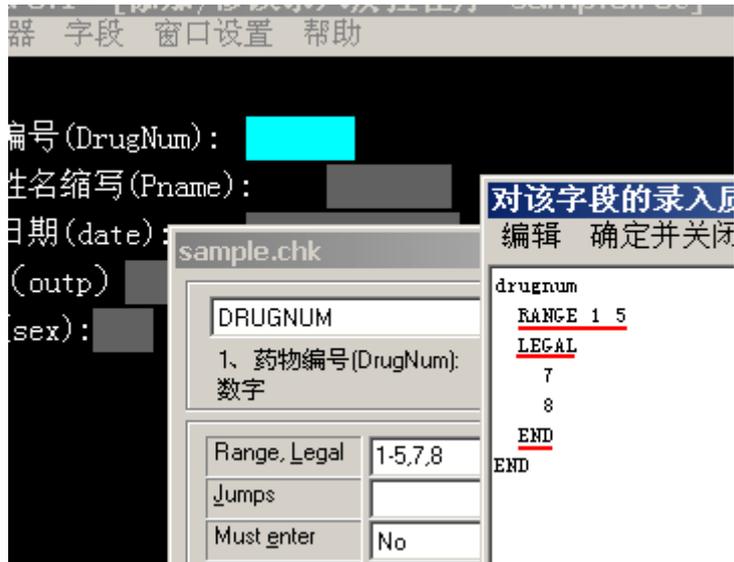


图 3-6 CHK 文件的“编辑”

这是对当前字段的 CHK 命令进行文本方式的编辑（其实 Epidata 的文件基本都是文本格式，只不过文件的后缀不是.txt 罢了）。在里面我们可以看到，一个合格的 CHK 命令是字段名加 CHK 命令，以 END 结尾。

Range 命令的用法是

RANGE 1 5 （它定义了允许的数值范围是 1 到 5）

Legal 的用法是：

LEGAL

1

5

END

（注意：很多 CHK 命令都以 END 结尾，比如这个 LEGAL，但 RANGE 不需要）

对于 LEGAL 命令，每个允许的数值要占一行。

事实上，对于这两个命令，我们不需要以文本方式编辑，直接在图 3-4 中的选项卡中填上数值范围或数值就可以了（以英文“,”分隔每个单独的数值），而对于选项卡中没有的命令，我们必需用编辑文本自己编写。

好了，这两个命令我们已经简单的了解了用法，下面我们看看跳转命令 Jumps 怎么用。

第 4 节 数据录入的控制(2)

2、跳转

对应的 CHK 命令：**Jumps**

“控制”作用：在录入时，如果录入的值符合条件，则跳转到条件中设定的字段处进行录入。

这个一个很常用的 CHK 命令，格式也很简单，如下图：



图 4-1 设置跳转命令

在上图中我们可以看到，如果在 sample.chk 那个选项卡中编辑 Jumps 命令的话，格式是：

Y>sex

解释：**如果录入的值等于 Y 则将跳转到字段 sex**。对于这个字段，由于是布尔型变量，录入 0 或者 1，系统自动将其转换为 N 或 Y，所以这个字段的 Jumps 命令中，条件取值是 Y 而不是 1。

如果进行文本格式的编辑（见上图中右下角的编辑窗口），Jumps 的命令格式是：

outp /这个是字段名

JUMPS /Jumps 命令开始

Y sex /跳转条件，可以有多个，每个条件占 1 行

END /Jumps 命令结束

END /outp 字段的 CHK 命令结束

上面这种情况，是一个跳转命令，如果有多个跳转怎么办？

在 CHK 选项卡中设置的话，是用英文的“,”分隔不同的跳转命令，如 Y>sex,N>inh。

这里有一个小窍门，在使用 CHK 选项卡的实际操作中，我们不必书写 Jumps 将要跳转到的字段名称，用鼠标就可以搞定：

当在 CHK 编辑选项卡中 Jumps 的编辑框中，已经输入了“Y>”之后，我们用鼠标点击数据录入界面（图 3-7 中黑色背景）中，我们需要跳转到的那个字段对应的输入框（黑色背景中的灰色输入框），比如我们点击“性别”后面那个输入框，则我们正在编辑的 Jumps 命令“Y>”将成为“Y>sex”，而不必我们手动输入了。

以上介绍了 Jumps 命令的常用方法，其它的跳转条件暂不在这里介绍。

3 是否必需录入

对应的 CHK 命令为：**MUSTENTER**

“控制”作用：在录入时，如果设置了 MUSTENTER（默认是 No，不设置），则该字段必需录入数据，即不能为空（当然对与错它是不管的）。设置见下图：

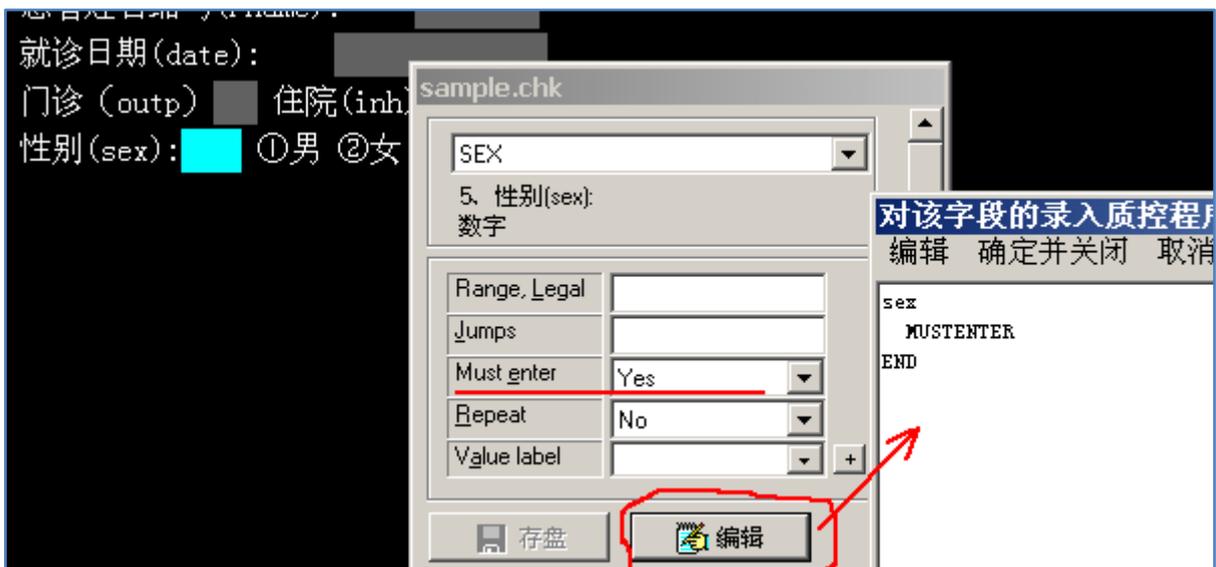


图 4-2 MUSTENTER 的设置

MUSTENTER 的设置非常简单，在 CHK 选项卡中，只有两个选项，第一个是“NO”，第二个是“YES”，上图中选择了“YES”。

点击 CHK 选项卡下方“编辑”按钮，用文本方式打开该字段的 CHK 命令，我们可以看到这个命令的格式：

Pname /字段名称

MUSTENTER /这是设置为“YES”的情况，设置为“NO”则 MUSTENTER 不出现。

END / 该字段的 CHK 命令结束，注意不是 MUSTENTER 的，MUSTENTER 是少数几个不需以 END 结尾的 CHK 命令。

4 是否允许重复值

对应的 CHK 命令为：**Repeat**

“控制”作用：在录入时，如果设置了 Repeat（默认是 No，不设置），则在录入时，**设置了 Repeat 的字段，将自动填充上一条记录中该字段的值。**当然这个值是允许被修改的。

这个命令主要用于字段值较少变化的情况，比如我们的调查表最后一项是观察医生的姓名，可能几百个病人的记录，只有几个观察医生，那么我们就可以应用 Repeat 命令，在我们录入数据时，如果当前数据中观察医生项有一个值，那么在我们录入下一条记录时，观察医生项还是这个值，不用我们去录入了。

如果当前记录，修改了这个字段的值，那么在之后录入的数据中，将调用修改后的新值。

这个功能非常实用。

Repeat 命令的格式与 MUSTENTER 是相同的，这里就不截图了。

其实，还有另外一种情况，那就是在一条记录中，如何重复当前记录中某一字段的值，这个实现起来也很简单，大家可以动动脑筋。

5 字段的标签

启用标签，对应的 CHK 命令为：COMMENT LEGAL USE （USE 后面跟标签的名称）

标签的“控制”作用：在录入时，如果设置了字段的标签，则在录入该字段数据时，按 F9 或者数字键盘上的 + 号，则该字段定义的标签将被调出来显示，以达到提示作用。如下图：

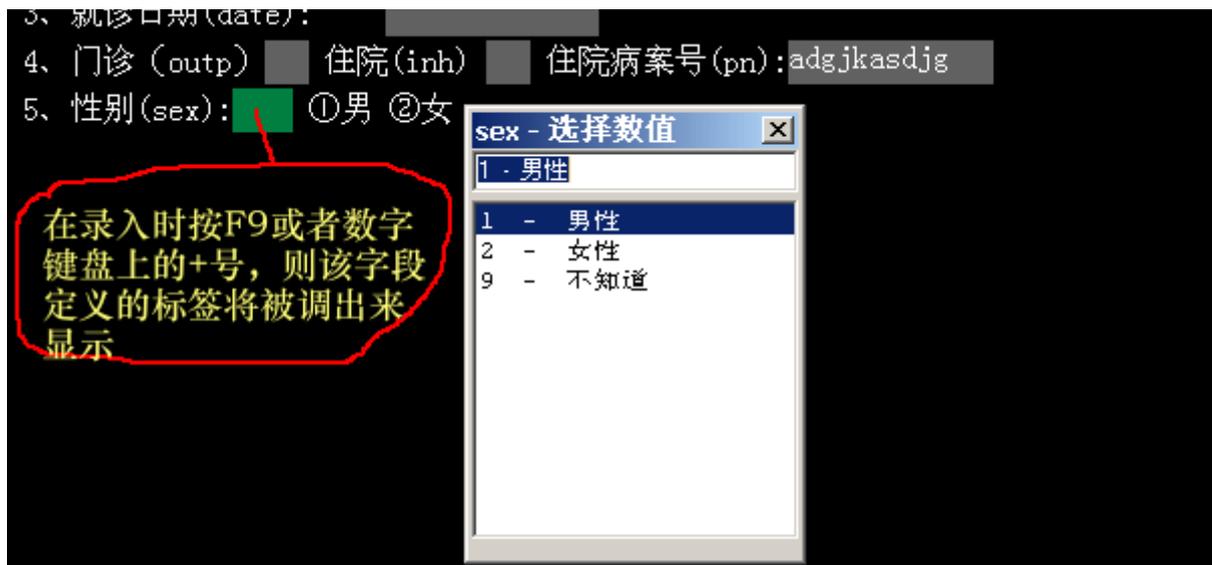


图 4-3 字段标签的提示作用

目前在 EpiData 系统中，已经定义的标签有 3 个，标签名称分别是：agegroups、continent 和 sex，分别用于年龄组、洲及性别的标示。

EpiData 可对现有标签进行编辑，也可自定义新的标签，方法有两种，一个是更改 epidata 目录中的 EpiData.lbl 这个文件，用记事本打开，按照其中的格式加入自己的标签就可以了，也可以在 CHK 文件编辑器里来自建，在图 4-2 中，Value Label 选项的右侧有一个“+”号，点击看看，肯定有收获。

我个人觉得标签这个功能并不实用, 如果我们在录入数据的时候, 还要按 F9 或者什么+号来看看我们需要录入什么数值, 那么录入效率肯定成问题。

一般做法是在字段后面直接把需要标示的内容写出来, 在数据录入时一目了然, 比如性别, 输入框后面已经明确 ①男 ②女, 在录入时根据情况录入 1 或者 2 就可以了, 所以事实上这个命令我从来没用过。

好了, 至此, 我们已经把 CHK 选项卡上列出的 5 个 CHK 命令都作了简单介绍, 相信大家已经能够写出符合自己要求的控制文件了。

当然, 如果要实现更为复杂的“控制”, 单纯用以上 5 个命令是无法完成的; 比如: 有字段 A、B、C、D, 都是数字型的, D 的值=A+B+C, 如何实现 D 值自动填充? 再比如, 录入了 A, B 后, 如果让 C 自动填充上 A 的值? 等等, 因为这个教程是教大家快速入门, 所以不可能介绍得更多, 有兴趣大家可多花时间看看 Epidata 的使用手册。

第 5 节 结语

到这里, 我们已经把在 epidata 中 如何建立调查表文件.qes, 如何生成数据库文件.rec, 如何编辑数据录入的控制文件.chk, 都作了介绍, 其它的如数据处理, 数据导出就不再写了, 本身这些都不难, 基本上都是菜单操作, 而且因为是入门性质的文章, 所以不可能对每个功能每个细节都讲到, 我觉得, 对于一个初学者, 如果读了这几篇入门文章后, 能够自己建库并完成数据录入, 最后将数据导出为自己需要的格式, 也就达到了目的。

Epidata 真的是一款非常不错的软件, 虽然它还很有很多不足之处, 但在目前的情形下, 它作为数据录入和管理的软件, 还是可以满足基本要求的。如果该软件在现有基础上, 开发、设计用户与权限、日志、数据修改痕迹记录等相关功能, 相信能够完全胜任临床试验对数据管理的严格要求。

对于 Epidata 的使用, 上手是非常容易的, 想要把它的功能运用得淋漓尽致, 恐怕这个 90 分钟的快速入门教程是远远不够的, 还是那句话, 多看看 Epidata 3.0 使用手册, 这本由北京大学公共卫生学院吕筠老师翻译的手册, 一百多页, 图文并茂, 是学习 Epidata 的必备资料, 下载地址:

http://www.epidata.dk/downloads/epdhelp_cn.pdf。

Medleaf 于 2009 年 8 月