

Stata——一个新型统计软件

王建民

由于我国计算机的应用起步较晚,目前国内尚无一个真正成熟的软件,因此近十几年从国外引进了多种统计软件,如SAS、BMDP、SPSS、GLIM、SYSTAT等。这些软件的引入,对我国卫生统计事业的发展起了很大的推动作用。国外统计软件的发展速度也很快,各软件都在逐步完善自己。通过克服本身的缺点,借其它软件的优点,同时还出现了一些新的统计软件。现将美国新出现的统计软件Stata 2.1的几个优点分述如下。

1. 计算速度快: Stata的所有计算都在内存中进行,不与磁盘交换数据, Stata可以使用计算机的所有内存。另外, Stata还可以利用计算机的协处理器。由于计算都在内存中进行,所以Stata能够处理的数据量大小,取决于计算机内存的大小。但由于计算机工业的迅速发展,就目前的情况来说,一般不会由于计算机内存太小而不能处理我们的数据。设我们的每个观察有50个变量,那么640K内存可处理2400例观察,1M内存可处理6000例,2M内存可处理12000例等。

2. 使用方便: Stata的每一个命令都对应于一个操作或运算,不象有些统计软件,需要几条语句才能完成一个操作或运算。例如,设数据中有两个变量, y 为因变量, x 为分组变量,现在想做单因素方差分析,那么在SAS中就要用下列命令: PROC ANOVA; CLASS X; MODEL Y=X; RUN; 而在Stata中,则只用一个命令,即: oneway y x

3. 图形功能强: Stata所产生的图形都是在图形模式下进行,而不是在文本模式下进行。并且产生的图形可在各种普通打印机、激光打印机和绘图仪上输出。还有一些新的图形,如散点图矩阵,它对应于相关矩阵,把相关系数替换成了散点图。使变量间的关系以图形形式给出,既可揭示线性关系,也可揭示非线性关系,同时使使用者对变量间的关系有一个直观影响。

4. 兼容性强: 软件间的数据交换功能是衡量一个软件好坏的重要标准。Stata能够读取ASCII码数据文件,也能够把Stata数据文件转换成ASCII码文件。现在流行的各种软件基本上都可读写ASCII码文件,因此Stata与其它软件的数据交换是没有问题的。

Stata还可把产生的图形转换成LOTUS的图形。

5. 统计内容充实: 包括简单的描述统计到较复杂的多因素方差分析、多元回归、逐步回归及各种非参数

统计方法。还有Logistic模型、Probit模型、Cox模型及其它生存分析方法。主要命令及功能如下: anova: 单因素、两因素及多因素方差分析和协方差分析; correlate: 计算Pearson相关系数; cox: 估计Cox比例风险模型; logit: 估计Logistic模型; probit: 估计Probit模型; oneway: 单因素方差分析及两两比较; predict: 各种模型的预测值、残差、t化残差、标准化残差等; regress: 各种线性回归; stepwise: 逐步线性回归; summarize: 产生描述性统计量,如均数、标准差等; tabulate: 产生行列表及分组描述性统计量; test: 对各种回归系数进行检验; ttest: 配对和成组t检验。

6. 数据管理功能: 它具有一些简单的数据管理功能,如数据的合并与连接,按条件、范围进行转换、删除,修改变量名,压缩数据,排序等。另外它还有一个特殊命令,即inspect,这个命令可对数值变量进行一个大致描述,使用户能够熟悉变量的一般情况。

7. 软件占用磁盘空间少: Stata软件只占1.8M的磁盘空间,而SAS要占用20多兆磁盘空间。

8. 可编程: 利用Stata的命令可以编程, Stata内部的一些命令如ttest就是用Stata命令编写的程序,并且这种程序是可读的。我们可以通过学习,分析这些程序,得到一些编程技巧和思想,并且利用Stata命令产生的一些结果,通过编程进行更深入的分析。

9. 易学、容易掌握: 说明书讲解详细,举例丰富,每个功能都配有例题。对于统计命令,还给出有关的公式及模型,此外还有详尽的帮助功能和菜单驱动功能。

1992年出现了Stata 3.0, 它的统计功能比Stata 2.1有很大的提高,增加的主要内容有流行病学调查中的各种专用四格表、寿命表、条件Logistic模型(可处理 $M_i : N_i$ 的数据)、多状态Logistic模型、有序Logistic模型、各种模型的逐步回归、Torbit模型等。

从Stata 3.0的情况来看, Stata充分考虑了卫生统计的要求,增加了很多流行病学的统计方法,在Logistic模型中就可让Stata输出OR值及OR的可信限。

总之, Stata软件是一个很适合于卫生统计专业的软件,基本上能够满足卫生统计的各种要求。

(收稿, 1993-7-20)