

Stata入门

Stata软件介绍

Stata是一个用于分析和管理数据的功能强大又小巧玲珑的实用统计分析软件，由美国计算机资源中心（Computer Resource Center）研制。它同时具有数据管理软件、统计分析软件、绘图软件、矩阵计算软件和程序语言的特点，又在许多方面别具一格。操作灵活、简单，易学易用。

Stata的统计功能很强，除了传统的统计分析方法外，还收集了近20年发展起来的新方法，如Cox比例风险回归，指数与Weibull回归，多类结果与有序结果的logistic回归，Poisson回归，负二项回归及广义负二项回归，随机效应模型等。具体说，Stata具有如下统计分析能力：数值变量资料的一般分析：参数估计，单因素和多因素的方差分析，协方差分析，交互效应模型，平衡和非平衡设计，嵌套设计，随机效应，多个均数的两两比较，缺项数据的处理，方差齐性检验，正态性检验等。

Stata16主要包括以下两方面的升级：与大数据相关的功能、继续深耕计量经济学的经典与前沿方法。

（1）与大数据相关的功能

Stata16在大数据方面提供了一些重要的新功能特性，极大方便了用户的使用。例如Python Integration提供了与Python的接口，让用户可以在Stata界面下调用Python，使用Python的各种机器学习方法，并在Stata中显示结果；import sas与import spss可以将SAS与SPSS的数据导入Stata；Meta-Analysis使得元分析变得方便、快捷与高效，并辅之以强大的可视化功能；Multiple Datasets in Memory使得Stata16可在内存中同时调用多达100个数据集；新的Lasso命令可以很方便的计算回归系数的整个路径。

（2）继续深耕计量经济学的经典与前沿方法

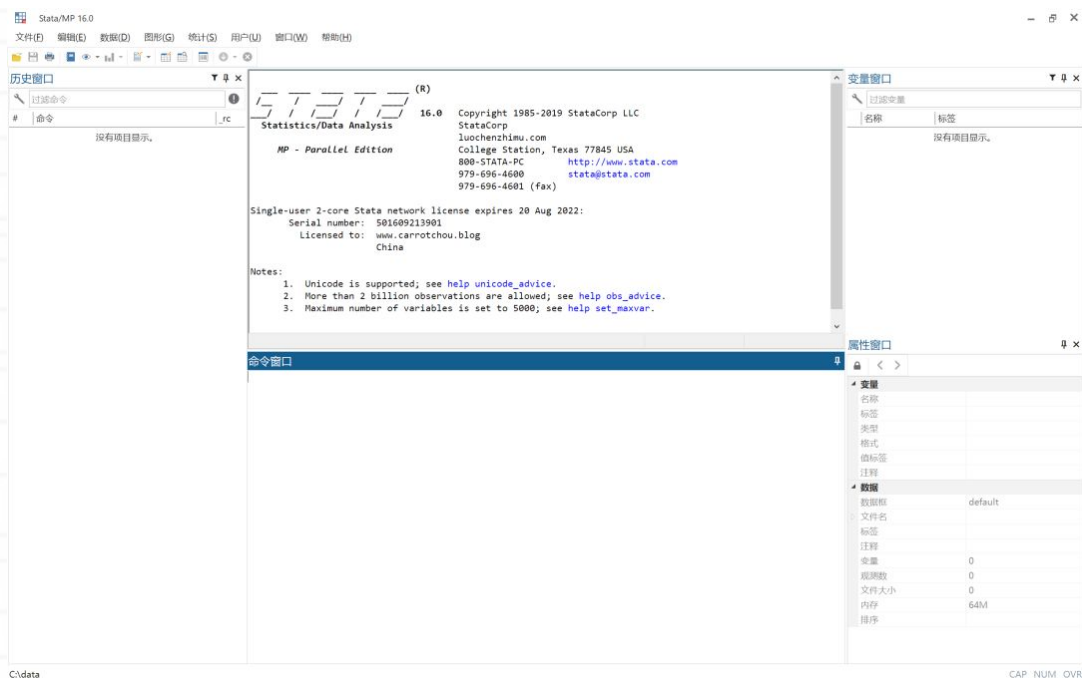
Stata16新的npregress series命令填补了Stata在非参数回归领域的空白，使得非参数序列回归变得方便而高效；对于微观计量中常用的离散选择模型，Stata16专门设立了Choice Models选择模型；Panel-data ERMS将ERMS推广到了面板数据中，新引入了xtegress、xteintreg、xteprobit、xteoprobit等强大命令；贝叶斯分析模块提供了一些新的功能特性。

——内容摘自上海财经大学教学实验软件平台

官网：<https://www.stata.com/>

Stata的窗口界面

#从 15.1 版本开始，官方为 Stata 提供了简体中文语言包，所以大家使用起来就更加方便了。



Stata的界面主体由五个部分构成

左上 “Review” (历史窗口)：记录着自启动 Stata 以来的命令。

中上 “Results” (结果窗口)：显示执行 Stata 命令后的输出结果。

中下 “Command” (命令窗口)：在此窗口输入 Stata 命令。

右上 “Variables” (变量窗口)：记录着 Stata 内存的所有变量。

右下 “Properties” (性质窗口)：显示当前数据文件与变量的性质。

#这五个窗口是可以根据自己的喜好自由移动，改变大小的。

实际使用中，我们是在 “Command” （命令窗口）中输入代码来实现具体操作单，这和R与pycharm有点像。

此外Stata还有一行菜单栏和一行快捷键栏



Stata的实例操作

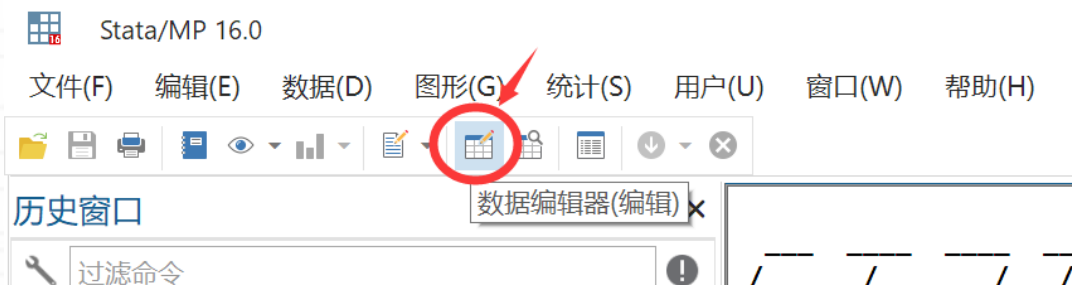
#数据均来自陈强教授的《高级计量经济学及Stata应用（第二版）》中的数据

以 Nerlove(1963)对电力行业规模报酬的经典研究来介绍 Stata的实际操作。该数据集 nerlove.xls(Excel 文件)包括了 1955 年美国145 家电力企业的横截面数据。

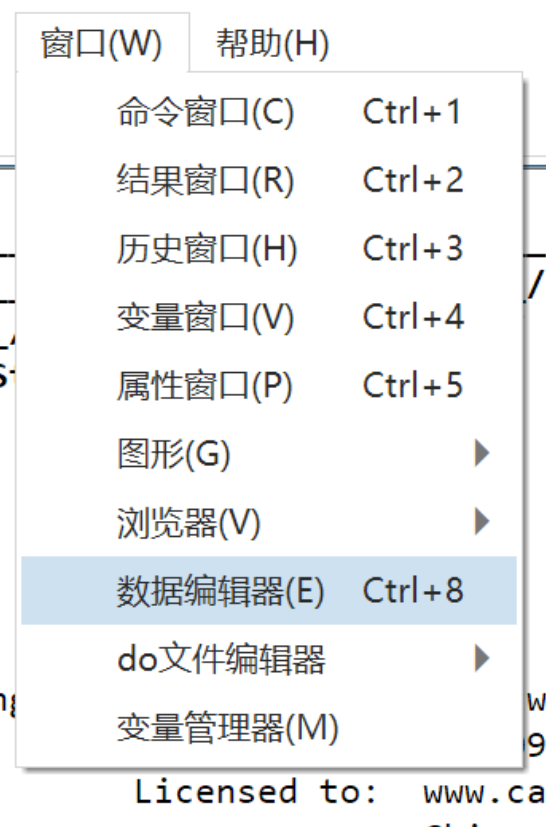
导入数据到Stata

Stata导入数据有两种方法，一种是通过打开 “数据编辑器” 将Excel的数据复制到Stata中。

数据编辑器的打开也有两种方法，一种是通过Stata界面的快捷键



另一种是通过点击菜单中的“窗口”→“数据编辑器”打开



另一种，将数据导入Stata的方法是通过菜单中的“文件”→“导入”（这种方法可以导入多种格式文件的数据，适合数据量非常大的时候使用）



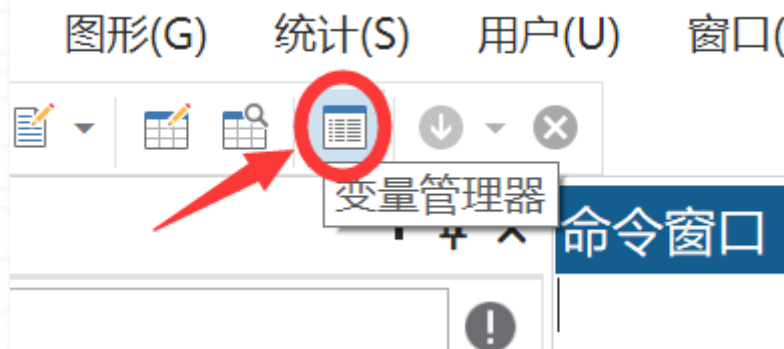
导入数据后，关闭数据编辑器界面，回到stata主界面可以看到在变量窗口中出现数据变量的信息

变量窗口		Y	Q	X
过滤变量				
名称	标签			
tc	total cost			
q	total output			
pl	price of labor			
pf	price of fuel			
pk	user cost of capital			

#数据均来自陈强教授的《高级计量经济学及Stata应用（第二版）》中的数据

数据中一共有五个变量，分别为 tc(total cost, 总成本), q(total output, 总产量), pl(price of labor, 小时工资率), pf(price of fuel, 燃料价格), 与 pk(user cost of capital, 资本的租赁价格)。

在“变量管理器”中可以更改变量的变量名、标签以及变量的存储格式，类似于spss的内容。



点击保存可以将数据保存为stata的格式，即“.dta”

可以通过命令**use“文件路径\文件名.dta”,clear**的方法打开，其中逗号“,”之后的“clear”为“选择项”(option)，表示可替代内存中的已有数据。

例如：使用use"E:\nerlove.dta"打开数据nerlove。

导入数据后我们就可以通过在命令窗口输入代码来实现相应的统计分析了。

