

Stata统计分析入门

中国人民大学统计学院

谢益辉

2007-04-20

中国人民银行Stata培训

培训提纲

1. Stata 软件特点
2. 界面和语言介绍
3. 数据操作与运算
4. 统计分析与展示

1. Stata 软件特点

- a) 自身优势（为什么要用Stata）
- b) 同类软件比较（SAS、SPSS等）
- c) 四个版本的比较

1.1 Stata的优势

□ 快速、准确、易用

- 既可以鼠标操作，亦可执行语句（Syntax）

□ 统计模型方法广泛

- 诸如基本描述统计、列联表、广义估计方程（GEE）、线性/广义线性模型（GLM）、（多变量）方差分析（ANOVA/MANOVA）、聚类分析等等

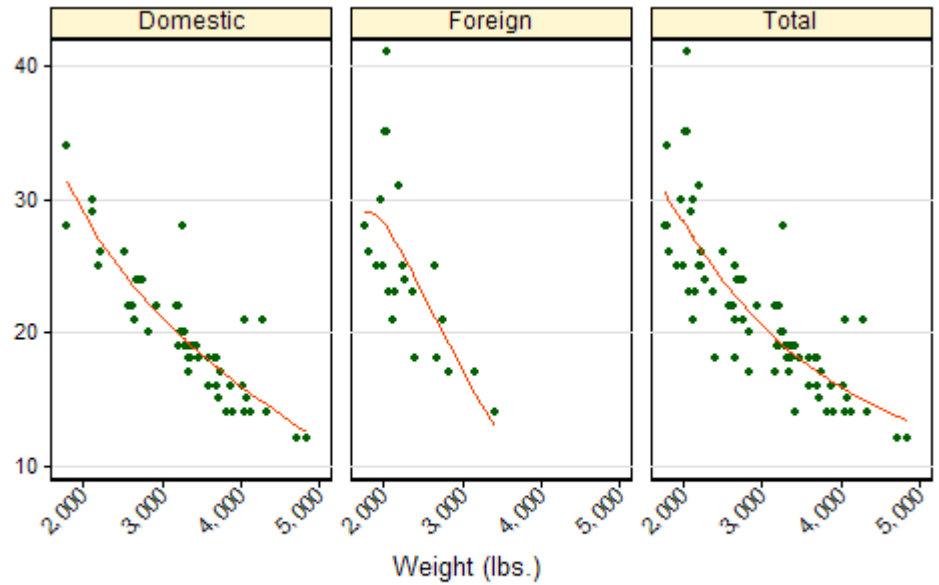
□ 数据管理功能齐全

- 各种类型计算机数据集、变量操作、统计数据

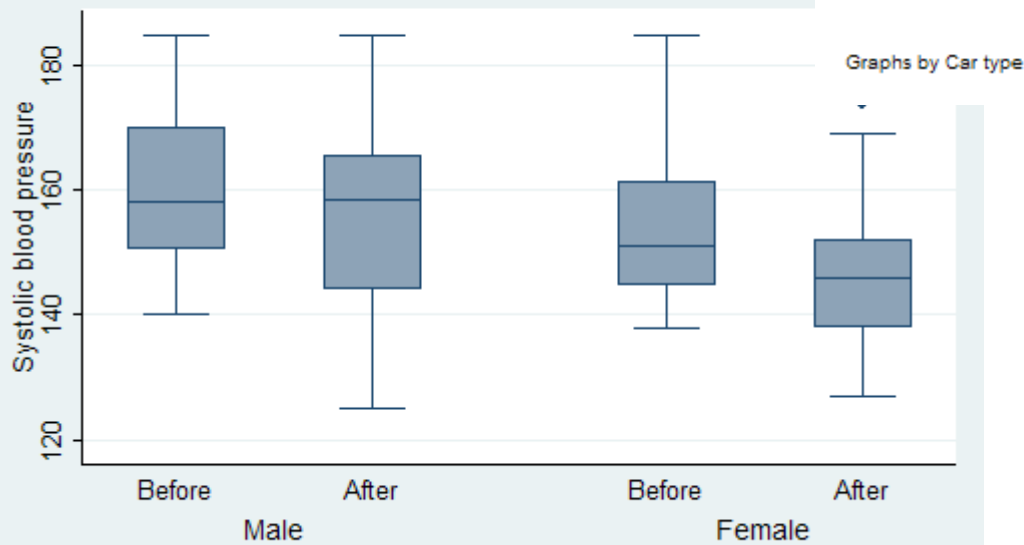
The screenshot displays the Stata/SE 9.0 interface. The **Statistics** menu is open, showing options like 'Summaries, tables, & tests', 'Linear models and related', and 'Time series'. The 'Time series' submenu is also open, highlighting 'Regression with Newey-West std. errors'. A search window titled '(#1) [search regress, all]' is active, showing search results for 'regress', including 'Official help files, FAQs, Examples, SJS, and STBs'. The Command window at the bottom contains the command `regress mpg weight`. The Variables window on the left lists variables such as 'make', 'price', 'mpg', and 'weight'.

1.1 Stata的优势（续）

- 高质量图形输出
 - 回归拟合、分布、时间序列、生存分析等
- 回应及时、可扩展性高
 - 具备编程的功能，可以任意扩展分析方法
- 矩阵编程
- 跨平台兼容
- 多种文档、出版物
- 技术支持、学习资源



Response to Treatment, by Sex
(120 Preoperative Patients)



Source: Fictional Drug Trial, StataCorp, 2003

1.1 Stata的优势（续）

- 广泛使用
- 费用具有可选择性
 - 对象：企业、教育机构、政府等
 - 形式：单用户、网络、批量、租约

1.2 与SPSS、SAS等软件比较

□ 怎样的软件叫强大？

- 功能齐全
- 编程方便
- 模块化
- 技术支持服务
- 小巧
- 透明

1.3 四个版本的比较

□ Stata SE

- 支持大数据集

The logo for Stata SE, featuring the text "STATA®/SE" in a bold, sans-serif font with a reflection effect below it.

□ Intercooled Stata

- Stata标准版本

The logo for Intercooled Stata, featuring the text "intercooled STATA®" in a bold, sans-serif font with a reflection effect below it.

□ Stata/MP

- 支持双核、多核/多处理器的计算机

The logo for Stata/MP, featuring the text "STATA®/MP" in a bold, sans-serif font with a reflection effect below it.

□ Small Stata

- 学生版（教育版）

The logo for Small Stata, featuring the text "small STATA®" in a bold, sans-serif font with a reflection effect below it.

1.3 四个版本的比较（续）

- Stata/SE与Stata/MP最多能处理有32,766个变量（列）的数据集，样本量的限制则取决于电脑内存。至多允许运算11,000 x 11,000的矩阵（也受内存限制），也就是说，这两个版本能处理的自变量的最大数目为10,998。
- Intercooled Stata允许处理至多2,047个变量，样本量同样也受内存限制，最多接受798个自变量。
- Small Stata最多处理99个变量、大约1,000个观测，一个模型中最多允许38个自变量。

1.3 四个版本的比较（续）

版本	变量数	自变量数	样本量	矩阵规模	并行处理	操作平台
Stata/MP	32,766	10,998	不限*	11,000	能	Windows, Macintosh或 Unix
Stata/SE	32,766	10,998	不限*	11,000	否	Windows, Macintosh或 Unix
Intercooled	2,047	798	不限*	800	否	Windows, Macintosh或 Unix
Small Stata	99	39	1,000	40	否	Windows或 Macintosh

* 只受计算机内存限制

1.3 四个版本的比较（续）

版本	内存下限	磁盘空间下限
Stata/MP	256 MB	60 MB
Stata/SE	256 MB	60 MB
Intercooled Stata	128 MB	60 MB
Small Stata	128 MB	60 MB

2. 界面和语言介绍

- 各个窗口说明
- Stata的程序语言
 - 语法
 - 函数

3. 数据操作与运算

- 数据输入与保存
- 数据的增删改

4. 统计分析与展示

- 一般描述统计与简单置信区间
- 数值资料分析
 - 统计检验
 - 方差分析/协方差分析
- 离散资料分析
 - 列联表独立性检验等
- 经典统计方法
- 图示功能