# 2017(秋)计量经济学- Stata 操作入门(II)

南京大学经济学系

凌晓光

2017年9月25日

**这是stata入门的第二讲**

**\*上节课要点回顾：**

1.stata狂霸酷拽，比别的软件不知高到哪里去了

2.stata有各种版本，想要丝滑体验，推荐14.0的MP版本

3.图形化界面初识：菜单栏，工具条，窗口。

4.do文件的编写规则初识

**\*本节要点：**

1.数据导入

2.log文件的使用

3.双样本t检验

4.常用的stata命令：help文件，安装程序包，数据描述，图表

## 数据导入

我们使用计量软件的目的是对“数据”施加“命令”，以得到结果。相应地，在Stata中，最主要的文件类型包括数据(.dta)和命令(.do)，下面我们分别介绍如何在Stata中对二者进行操作。

(1) Stata所直接处理的是扩展名为.dta文件，类似txt文档，占用存储空间小

\*可以在菜单栏打开

clear all

use "C:\Users\Xiaoguang\OneDrive\2017计量经济学\9.18计量经济学-stata入门/rawdata/Training.dta"

(2)其他兼容的数据类型 csv,txt, xlsx

clear

insheet using Training.csv

clear

insheet using Training.txt

xlsx文件(stata自身具有一定的数据格式转换功能)

clear

import excel Training.xlsx, sheet("training") firstrow //第一行为变量名

(3)复制粘贴

(4)stat transfer

stat/transfer 12.0 似乎不能在win10 1703版上正常运行，请大家试一下并反馈是否出现“a debug report has been generated”这样的报错；我装了9.0版本，目前正常运行

数据的保存

save "C:\Users\Xiaoguang\OneDrive\2017计量经济学\9.18计量经济学-stata入门/workingdata/Training\_cleaned",replace

## 2. 录屏神器：log文件

log using "C:\Users\Xiaoguang\OneDrive\2017计量经济学\9.18计量经济学-stata入门/log/矩阵加法" //开始录制

matrix input a = (1,2\3,4)

matrix list a

matrix input b = (1,1\1,1)

matrix list b

log off //暂停录制

matrix c=a+b

log on //继续录制

matrix list c

log close //结束录制

## 3. 双样本均值t检验（别急，这个命令的句法在后面讲）

clear all

use "C:\Users\Xiaoguang\OneDrive\2017计量经济学\9.18计量经济学-stata入门/rawdata/Training.dta"

Ho:培训前，处理组和控制组收入均值无差异

ttest re74 , by(train)

Ho:培训后，处理组和控制组收入均值无差异

ttest re78 , by(train)

## 4. 常用命令

**4.1 help文档**

help ttest

**如何阅读help文档**

关于保存值

return list

**4.2 程序包安装**

为什么stata可以识别ttest命令？

Stata的程序包：标配程序包+不断发展的各种程序包

如何安装新程序包

**4.3 简单的数据描述**

clear all // 清空数据、变量

set more off , perm //关闭more功能,该功能在15.0之前默认是on的，请关闭它

global root "C:\Users\Xiaoguang\OneDrive\2017计量经济学\9.18计量经济学-stata入门"

/\* 利用全局宏变量设置根目录，每次开关stata软甲，全局宏变量都会被清除，

否则全局宏适用于全篇\*/

cd "$root/rawdata" //设置工作路径，以后调用.dta数据只需要 use filename 即可

use Training, clear //调取数据文件

des //描述

sum //摘要

tab mostrn train //列联表

collapse re74 re75 re78 ,by(train) //按照train分组并保留均值

list //列示每个样本的各个变量值

**4.4 scatter, graph pie**

Stata的图表有很多种，大家具体使用到时可以参考工具书查找需要的命令；

利用命令修改图表比较复杂，也可以在图表编辑器里进行编辑。

**4.5 输出estpost esttab**

这两个命令的option非常多，大家有兴趣可以研究一下，真的非常多。。。。。

## 5. 一点微小的建议

(1)文件夹：分类保存不同类型的文件

尽量使用英文做文件夹和文件名(Stata13以后对中文支持变强)

举例：RawData,Dofile,Logfile,

也可以避免污染原始数据

(2)习惯利用do文件进行操作：编辑、修改更容易，操作可保存、复制

(3)定义宏变量：简化命令，方便修改

(4)充分利用搜索引擎Google(baidu很弱)、人大经济论坛、help文档等资源

(5)做好数据备份：Dropbox、Onedrive、百度云等网盘，本地介质，甚至邮箱

(6)善用help 文档，命令是 help+命令名，如 help ttest

## 6. 第一次作业-stata部分

**黑人悲催的故事**

假设在样本数据Training.dta中，低收入者既有白人(black==0)，也有黑人(black==1),他们中的一部分人在1975年被随机抽取并接受长期培训(train==1)。所有样本在1974年(培训前)和1978年(培训后3年)的收入分别记为向量re74、re78.

**问题：**

1.阅读 ttest 的帮助文档 (提示，命令是： help ttest, 注意其中关于判断语句 if 的说明)

2.利用 ttest 检验：对于全体样本，在1974年(培训前)，黑人和白人的收入均值有差异吗？1978年呢(培训后3年)？

3.利用 ttest 检验: 对于接受培训的样本(train==1),在1978年(培训后3年)，黑人和白人的收入均值有差异吗？对于没有接受培训的样本(train==0)呢？(提示：使用if语句)

4.利用estpost和esttab将问题2、3的t检验结果输出为txt格式的表格。问题2和3的对比说明了什么问题？

5.在变量窗口，我们看到变量re78对应的label是“real earns., 1978, $1000s”， 请将这个变量标签修改为中文“1978年的实际收入（千美元）”。 (提示：自行help label并阅读帮助文档)

**要求【重要，影响分数】：**

1.10月8日前，将do文件和log文件发送到邮箱：jl2017f@126.com (计量2017fall)

2.do文件和log文件均以“学号\_姓名\_专业”命名，如 b151090001\_户纳东\_经济.do

3.此次作业两个文件【不】要打包发送

**4.【重要】邮件主题请务必填写为“ 10月8日第一次作业 ”**

10月8日0:00之后邮件不会正确归档，请大家不要延误提交，错误的文件命名、主题或打包会妨碍批改，也会酌情扣分。