

连享会·直播课

Stata学习经验分享：从入门到进阶

龙志能 (上海财经大学)

longzhineng@foxmail.com



说在前面

- 基础分享 VS 进阶分享
- 概念框架 VS 具体操作
- 听了得学 VS 看就会了
- 个人经验，批判学习

分享目录

- Stata 入门的学习路线**总结**
- Stata 入门的学习资源**分享**
- Stata 进阶的学习路线**探讨**
- Stata 进阶的学习资源**推荐**
- 如何解决学习遇到的**难题**

Stata 入门的学习路线总结【系统学习，统计推断】

初级计量基础

- 多元线性回归 (OLS)
 - $Y = \beta_1 + \beta_2 X_1 + \dots + \beta_k X_k + \epsilon$
 - BLUE (高斯-马尔科夫定理; 最优线性无偏估计量)
 - 控制变量: 系数可能是不重要的, 但**控制变量选取**非常重要
 - 同方差假设, 零条件均值 $E(\epsilon|X) = 0$
- 虚拟变量与交互项
 - 构成关键变量 (高管性别, 企业违规)
 - 异质性分析 (分组回归, 交互项回归)
 - 虚拟变量陷阱 (i.类别变量; i.year i.industry)

初级计量基础（续）

- 模型正确选择
 - 离散型因变量
 - 二元变量 (logit)
 - 多元无序变量 (多元logit, mlogit)
 - 多元排序变量 (排序logit, ologit)
 - 受限因变量 (截堵, 企业捐赠, tobit模型)
 - 理论参考 + 文献参考 --> 最终选择
- 面板数据固定效应
 - 截面数据, 面板数据, 混合截面数据
 - 个体固定效应: 控制个体无法观测的不随时间变化因素
 - 双向固定效应: 同时控制**个体效应**+**时间**固定效应

Stata基础学习

- 基本数据处理（导入、合并、清洗、缺失值、异常值、重复值）
- 描述性统计与模型回归
 - 描述性统计
 - 相关系数分析
 - 组间差异检验
 - 模型回归（`reg, logit, tobit`）
- 固定效应模型回归(fixed effect)
 - 个体固定效应：`xtreg y x controls, fe robust`
 - 双向固定效应：`xtreg y x controls i.year, fe robust`

学习小tips【打开Stata】

变量命名推荐用英文

Tip1: Stata 文件路径

- `cd "D:\stata16\ado\personal\culture_risk"` // cd 路径设定
- `pwd` // 屏幕显示当前路径
- `dir` // 显示当前路径下文件
- `sysdir` // 显示系统文件夹
- `adopath` // ado程序存储位置([聊聊Stata中的profile文件](#))
- `cdout` // 打开路径所在文件夹

Tip2: do文档与实证文件夹

- 推荐使用** do 文档**记录命令，而非使用命令窗口和菜单窗口
- 新建 personal 文件（位于 ado 文件下）
- 实证文件夹建立在 personal 文件夹下，建议英文名
- 其内可分为：data(存放数据)；out(输出结果)；refs(参考文献)
- `cd "D:\stata16\ado\personal\culture_risk\data "`

Tip3: 善用help + findit（前者了解命令；后者模糊查询）

- `help xtreg` // `help + command` 针对具体命令求助
- `help language` // 通用语法格式
- `findit fixed effects panel` // 模糊查询
- `help xt`

Stata 入门的学习资源分享

书籍手册

- [Stata 16 manual](#)
- 计量经济学导论（伍德里奇）
- 《高级计量经济学及 Stata 应用》（陈强）

推文链接

- [连享会主页-推文列表-按类别](#)
- [Stata:实证结果输出命令大比拼](#)
- [Logit模型一文读懂](#)；[多元 Logit 模型详解 \(mlogit\)](#)；[一文读懂 Tobit 模型](#)
- [Stata: 面板数据模型一文读懂](#)

视频课程

- [连玉君-Stata 33 讲 --> 课件](#)
- [连玉君-直击面板数据 --> 课件](#)
- [连享会：数据清洗之实战操作\(一\)](#)
- [计量经济学及 Stata 应用](#)

Stata 进阶的学习路线探讨【干中学，因果推断】

中高级计量基础【适用条件+正确理解】

- 内生性问题理解
 - 审稿时，我乐于问及；投稿时，我怕被问及
 - 干扰项和解释变量相关；OLS估计结果有偏
 - 经验结果存在多种可能的解释 (并非“因果”推断)
 - 遗漏变量；衡量偏误；反向因果
- 随机化试验(Randomized Trials,RT)
- 匹配方法 (Matching)
- 工具变量法 (Instrumental Variable,IV)
- 双重差分方法 (Difference In Difference,DID)
- 断点回归 (Regression Discontinuity Design,RDD)

Stata进阶学习

- 高级数据处理
 - 处理复杂数据
 - Python 和 Stata 融合
- 构建重要变量
 - 模型计算：盈余管理，股价崩盘，过度投资
- 使用"**因果推断方法**"（按需）
 - 正确使用命令 + 适用条件检验 + 安慰剂检验
- 学习优秀的代码
 - 搜索网络代码
 - 参加培训课程
 - 期刊数据开放

Stata 进阶的学习资源推荐

书籍手册

- [《Endogeneity in Empirical Corporate Finance》](#)
- 《高级计量经济学及 Stata 应用》（陈强）
- 《因果推断实用计量方法》（邱嘉平）
- [《Mostly Harmless Econometrics》代码实现 \(基本无害的计量经济学\)](#)

推文链接

- [Stata专栏|内生性处理的常见方法;连老师PPT](#)
- [残差是个宝：盈余管理、过度投资、超额收益怎么算？](#)
- [稳健性检验！稳健性检验！](#)

视频课程

- [社会科学中的因果推论\(陈硕,复旦大学\)](#)
- [Mastering Mostly Harmless Econometrics](#)
- [连享会：数据清洗实战操作（二）](#)
- [连享会：因果推断专题](#)

小收藏

- [连享会 - 人文社科开放数据库汇总](#)
- [IRE 公开数据](#)
- [连享会：论文重现网站大全](#)

如何解决学习中遇到的难题

你遇到的 80% 的问题，早就被人回答过，你只要搜索（Search）就好。剩下的 20%，你才需要研究（Re-search）

搜索心法： 关键词选取、调整与组合

1、搜索引擎【面广但杂】

- 百度: baidu.com
- 必应: cn.bing.com
- 谷歌
- 举例:
 - 交互项做调节效应，调节变量和交互项显著，核心变量不显著可以吗？
 - FD-GMM和SYS-GMM都是用于动态面板吗？这俩有啥区别？

2、垂直网站【小而美】

- 连享会主页：
 - lianxh.cn
 - 连享会：Stata里看推文
- 经管之家：<https://bbs.pinggu.org>
- Statalist：www.statalist.org/forums
- Stack Overflow：<https://stackoverflow.com>
- 举例：
 - 如何输出实证回归结果？
 - 固定效应中，组内去心后常数项含义？
 - 如何解读回归系数的经济含义 economic significance

3、FAQs

- [Stata FAQs](#)
- 连享会 FAQs

4、移动终端

- [微信搜索](#)
- 连享会问答群

5、关于提问

- Search first! Search first! Search first!
- 如果是操作问题：关注**红色报错提示**(help error)
- 提问要**具体**，描述清楚问题**背景**，以及你的**努力**
- 如果可以，尽量用 Stata 中的 **dataex** 命令

说在最后

- 收藏 != 学会
- 实际操作 + 用于论文 + 分享讲解 = 真会了
- 运用Stata, 不要过于相信 Stata, Stata 也会撒谎!!!

连享会·主页 || 连享会·直播间 || 知乎推文



微信扫描，关注 “Stata连享会”



哔哩哔哩搜索 “小志小视界”

