Reproducible Analysis and Dynamic Documents Using Stata

Introduction:
Validity of science is bound to its reproducibility and thus any scientific finding is only valuable by assuming it is reproducible. Ideally, automating the process of data analysis, from the time that the data is digitized and stored on a computer to the point where all the analysis is carried out and the analysis report is written can significantly improve the reproducibility of the analysis.

Stata provides advanced tools for automating and carrying out sequences of data preparation and data analysis by nesting dofiles. However, it provides no tool for combining outputs, graphs, and text to produce the analysis report. As a result, adding graphs to Stata outputs and commenting and styling them have been a manual work for Stata users, which is slow, laborious, prone to human errors, and unreproducible.

In the workshop, I will discuss three user-written packages that complete the process of automating reproducible analysis for Stata users. MarkDoc, Weaver, and Ketchup packages support several markup languages such as Markdown, HTML, LaTeX, and AMC (Additional Markup Codes) that can be used to comment the analysis and generate an automated dynamic document directly from Stata Dofile-Editor.

The workshop only assumes basic familiarity with Stata commands and thus can be useful for a wide range of audience who use Stata on a daily basis.

The workshop would be particularly interesting for:
- Researchers at any level (student, post-doctoral fellow, research scientist, faculty) who wish to document the process of data analysis and create a reproducible analysis report within Stata
- Stata users who collaborate internationally and need to share an analysis report via internet
- Lecturers who teach statistics workshops using Stata and need to create workshop handouts
- Lecturers/researchers who produce online teaching materials about Stata and require HTML-based documents with syntax highlighter
- Researchers who wish to publish in journals that require an analysis script file and raw data to reproduce the analysis

For those who are curious to know more about MarkDoc, Weaver, and Ketchup packages prior to the workshop, I recommend visiting http://haghish.com/stata where I write about reproducible research and dynamic documents in Stata.

Participants should bring their own laptop with Stata 11, 12 or 13.

Date:
September the 1st 2015

Time:
12:15 – 16:00 2 hours introduction + 2 hours lab session

Location:
Auditory at AIAS (room 1632.201)

Teacher:
PhD student in Applied Statistics E. F. Haghish