Automatic generation of personalized answers to a problem set using \LaTeX \ & Stata

Stata conference - New Orleans
July 18 - 19, 2013

Rodrigo Taborda
r.taborda@uniandes.edu.co

School of Management
Universidad de los Andes - Bogotá
Introduction

- An econometrics or statistics course implies some sort of applied data analysis.
- Students are asked to use a computer software, data and report the results.
- The main purpose is to make students review concepts and procedures, inference and interpretation.
- Also, to introduce students to a real-life scenario in the analysis.
- Several books and websites provide databases for these purposes.
- Many of us write down problem sets along these lines: 1 problem set, $n$ students, 1 database, 1 (correct) answer.
- However, there are issues with this approach.
Issue: 1 problem set, \( n \) students, 1 database, 1 (correct) answer

- On the good side.
  - Obtaining the right answer is taken as a sensible work and understanding.
  - Reinforces the student who did the work.
  - There is a unique right answer.
  - Ease of grading.

- On the bad side.
  - Copy and paste the code or simply the output with answers an analysis.
  - One student doing the right work is enough for cheaters to take advantage (copy and paste).
  - When the database is popular (book or website) and the question is the same, this issue takes a world wide dimension.
  - Regression analysis in textbooks is almost always in line with theory and assumptions, far from a real-life analysis.
- So we end up with:
  - A pile of problem set answers on our desk.
  - With same answer and (probably) same analysis.
  - An answer increasingly far from any real life scenario, as assumptions and procedures work just fine.

- At undergraduate level this happening for a class of 50+ might be discouraging.

- The whole purpose of the assessment vanishes.

- When I have asked students use their own database, all kind of problems arise.

- Did the student learned or not?

- At graduate level, replication of published papers might solve some of these problems, however the issues around the 1 problem set, \( n \) students, 1 database, 1 (correct) answer remain.
Solution: 1 problem set, $n$ students, $n$ databases, $n$ answers

- Combine the use of Stata and \LaTeX.
- Take advantage that both generate an outcome out of commands.
- Stata uses: `.do`.
- \LaTeX uses: `.tex`.
- Exploit the ability of Stata of writing a \LaTeX file including Stata output.
Gini and Pasquini (2006) clearly describe how to use Stata to write a \texttt{\LaTeX} file, and then a \texttt{.pdf} file, including Stata output.

\textit{Automatic generation of documents}

Rosa Gini  
Regional Agency for Public Health of Tuscany  
Florence, Italy  
rosa.gini@arsanita.toscana.it

Jacopo Pasquini  
Regional Agency for Public Health of Tuscany  
Florence, Italy  
jacopo.pasquini@arsanita.toscana.it
How?

1. Write a `.do` file that:
   1. Resample the data for each student.
   2. Solve the problem set for each student.
   3. Captures Stata output in `local`, `graph`, `outreg2`, `sutex`, etc.
   4. Writes a `.tex` file with the answer for each student.

2. All students get a different data, one gets the “good” data, one gets the “bad”, one gets the “ugly”.
On the good side

- A unique and correct answer for each student. No cheating on the output and analysis.
- We will know the answer beforehand for each student data.
- Reinforces the work done by a honest student.
- Students will use an imperfect database, allowing them to think deeper into the regression results, and how to solve the usual issues.
- Zero mistakes in reporting result.
- Zero effort in generating 2, 20 or 200 problem set answers.
- You only have to write down the code once. Course by course, semester by semester, the code will rewrite $n$ personalized problem set answers to $n$ students.
**On the bad side**

- Introduce \textit{\LaTeX} and Stata to each other, nurture a good relationship and get them married sooner than latter.
- Control, care and fine tuning of the .do and .tex files.
- Does this really make students increase their effort, learning or understanding?
- Will not work for real time series data.
- A pile of problem set answers on our desk.
My experience so far . . .

- I have done this with cross section data (introductory econometrics) and panel data (intermediate econometrics), for moderate class size (±30 students).

- Three problem sets along the semester:
  - 1st PS asks description of data and graphs.
  - 2nd PS regression results and interpretation.
  - 3rd PS tests, diagnostics, assumptions violation.

- Less copy and paste.

- More analysis and attention to unusual data and results, to which I already have the answer.

- When grading, I can quickly check if the student’s answer is equal to mine, or not. Then I can concentrate in the student’s interpretation and computer-off work.
Example

/*DEFINE PATH*/
   cd C:\rodrigo\project_1st_latex_stata_text;
/*LOAD DATA*/;
/*EXTRACT 5% OF MOTHER DATABASE*/;
/*FOR EACH STUDENT*/;
   local student "Pedro Pablo"
   foreach student in 'student' {
      sysuse nlsw88, clear;
      sample 5;
      save data\nlsw88_`student'.dta, replace;
   }
/*BEGIN LATEX DOCUMENT*/;
   local student "Pedro Pablo"
   foreach student in 'student' {
      file open reporte_`student' using reporte_`student'.tex, write replace;
      file write reporte_`student' "\documentclass{article}\n;" _n;
      file write reporte_`student' "\usepackage{graphicx}\n;" _n;
      file write reporte_`student' "\setlength{\voffset}{-1in}\n;" _n;
      file write reporte_`student' "\setlength{\textheight}{24.2cm}\n;" _n;
Example

/*WRITE LATEX DOCUMENT*/;

file write report_‘student’ ‘"\begin{centering}" n n;
file write report_‘student’ ‘"\textbf{Name:}" student “ n n;
file write report_‘student’ ‘"\textbf{Answer to problem set. Introductory econometrics}" n n;

file write report_‘student’ ‘"\end{centering}" n n;
file write report_‘student’ ‘” n n;
file write report_‘student’ ‘” n n;
file write report_‘student’ ‘"\hrulefill " n n;

/*USE DATABASE*/;

use data\nlsw88_‘student’.dta, clear;

/*QUESTION 1*/;

/*DESCRIPTIVE STATISTICS*/;
sutex age wage hours ttl_exp tenure, labels

minmax
title(Descriptive statistics")
placement(h!)
key(tab:des_‘student’)
file(des_‘student’.tex) replace;

file write report_‘student’ ‘"\noindent \textbf{Question 1.}Descriptive statistics. "” n n;
file write report_‘student’ ‘"\input{C:/rodrigo/project_lst_latex_stata_text/des_‘student’.tex}" n n;

/*END LATEX DOCUMENT*/;

file write report_‘student’ ‘"\end{document}" n n;
file close report_‘student’;
Example

```latex
/*COMPILE LATEX*/;
    shell pdflatex report_`student`.tex;
    shell bibtex report_`student`.tex;
    shell pdflatex report_`student`.tex;
    shell pdflatex report_`student`.tex;
/*OPEN FILE*/;
    winexec "C:\Program Files\Adobe\Reader 10.0\Reader\AcroRd32.exe"
        "'C:\rodrigo\project_lst_latex_stata_text\reporte_`estudiante`.pdf'";
```