

Mass production of Reports in Stata

Karl Keesman

2009

3rd Australian and New Zealand Stata Users Group Meeting

The Problem

Publication quality reports can take lot of time to produce especially where large numbers of tables and other information needs to be added

The Problem

The Australian Early Development Index has about 600 report each with 8 tables and a lot of information between text

An example of the changing information:

8.8 % of the children are considered developmentally vulnerable. The highest proportion of developmentally vulnerable children are in the Local **TOWN X** (27.3 %) and the lowest proportion are in **TOWN V** (0.0 %)

There are 29 % of children performing well. The highest proportion of children performing well are in **TOWN F** (42.2 %) and lowest proportion are in **TOWN Z** (12.1 %)

The Problem

The Australian Early Development Index has about 600 report each with 8 tables and a lot of information between text

An example of the changing information:

8.8 % of the children are considered developmentally vulnerable. The highest proportion of developmentally vulnerable children are in the Local **TOWN X** (27.3 %) and the lowest proportion are in **TOWN V** (0.0 %)

There are 29 % of children performing well. The highest proportion of children performing well are in **TOWN F** (42.2 %) and lowest proportion are in **TOWN Z** (12.1 %)

A Solution?

There are some excellent user written programs: For example:

about

about is a table building program for oneway and twoway tables of frequencies and percentages, and for summary tables.

The output format can be csv, txt, html or LaTeX.

<http://www.ianwatson.com.au/stata/about8examples.pdf>

A Solution?

An example of tabout table:

Table 11: Average fuel consumption by origin

	Manufacturing origin		Total (mean mpg)
	Domestic origin (mean mpg)	Foreign origin (mean mpg)	
Category of gear ratio			
Low	16.0	.	16.0
Middle	20.6	17.0	20.4
Upper	21.1	23.5	22.1
Highest	27.8	24.7	25.5
Total	20.0	23.6	21.0
Weight category			
Low	28.7	26.7	27.4
Medium	20.5	17.0	19.8
High	15.8	.	15.8
Total	20.0	23.6	21.0
Roominess of the vehicle			
Minimal headroom	23.4	23.0	23.4
Adequate headroom	20.2	23.6	21.4
Excellent headroom	18.1	.	18.1
Total	20.0	23.6	21.0

Source: Stata Dataset. Population: All cars.

A Solution?

estout

estout, estab etc. has a huge range of options for regression table, one way table, summarise out put etc.

The output format can be csv, rtf, txt, html or LaTeX.

<http://repec.org/bocode/e/estout/>

A Solution?

Table: Regression table

	(1)	(2)
	Price	Price
Weight (lbs.)	1.747 (2.72)	3.465 (5.49)
Mileage (mpg)	-49.51 (-0.57)	21.85 (0.29)
Car type		3673.1 (5.37)
Constant	1946.1 (0.54)	-5853.7 (-1.73)
Observations	74	74

t statistics in parentheses

A Solution?

Where the text and the analysis code are in the same document:

weave

A Solution?

Mass producing appendices using Stata and word processor mail merge

NASUG July 11, 2005

by: Michael Blasnik

M. Blasnik & Associates, Boston, MA

The data format for a mail merge

Identifier
data

The data format for a mail merge

Identifier
data

table1 _r 1 _c 1	table1 _r 1 _c 2	table1 _r 1 _c 3	table1 _r 1 _c 41	table5 _r 2 _c 1	table1 _r 2
1	stuff	stuff	stuff	stuff	stuff

The data format for a mail merge

Identifier
data

table1 _r 1 _c 1	table1 _r 1 _c 2	table1 _r 1 _c 3	table1 _r 1 _c 41	table5 _r 2 _c 1	table1 _r 2
1	stuff	stuff	stuff	stuff	stuff

stvar _{table} ~1	stvar _{table} ~2	table _{level} ~1	table _{level} ~2	table _{level} ~3	table _{level} ~4	table _{level} ~5
Domestic	Foreign	1	2	3	4	

Clip from Stata data editor

To do this I will use Stata and Mata:

We could have used postfile but I think Mata is the logical choice, because:

- 1 Easy to manipulate matrices

To do this I will use Stata and Mata:

We could have used postfile but I think Mata is the logical choice, because:

- 1 Easy to manipulate matrices
- 2 Mata has lots of functions that allow you to do almost anything to the data eg. finding the max value

To do this I will use Stata and Mata:

We could have used postfile but I think Mata is the logical choice, because:

- 1 Easy to manipulate matrices
- 2 Mata has lots of functions that allow you to do almost anything to the data eg. finding the max value
- 3 Mata is fast

To do this I will use Stata and Mata:

We could have used postfile but I think Mata is the logical choice, because:

- 1 Easy to manipulate matrices
- 2 Mata has lots of functions that allow you to do almost anything to the data eg. finding the max value
- 3 Mata is fast

An example

Today we will look at how these are produced.

Assume that we wish to produce the following report in Word

Car Repair Report

It was found that highest category had cases XXXX and this was for XXXX cars with XXXX repairs per year.

	No of Repairs in 1978				
	1	2	3	4	5
Foreign	2	8	27	9	2
Domestic	0	0	3	9	9

Assume that we wish to produce the following report in Word

Car Repair Report

It was found that highest category had cases **XXXX** and this was for **XXXX** cars with **XXXX** repairs per year.

	No of Repairs in 1978				
	1	2	3	4	5
Foreign	2	8	27	9	2
Domestic	0	0	3	9	9

The example in Stata