

Regression Presentations: Tables

Paul E. Johnson¹ ²

¹Department of Political Science

²Center for Research Methods and Data Analysis, University of Kansas

2013

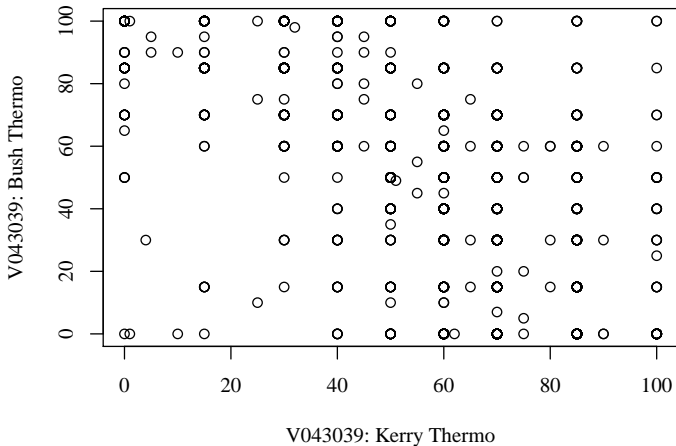
Presenting OLS To The Masses

- We Nice Looking Regression Tables
- Better to output a close-to-final result from R
 - Reduces typographical errors
 - Easy to re-run estimates and produce another automatic table.

Get Some Data from Nat. Election Study 2004

```
## V043038      B1a. Feeling Thermometer: GW Bush
## V043039      B1b. Feeling Thermometer: John Kerry
## V043210      R1. R position on gay marriage
## V043213      S3. National economy better/worse since
                GW Bush took ofc
## V045117      G4a. Liberal/conservative 7-point scale:
                self-placement
## V045145X     H5x. Summary: Pre-Post US flag makes R
                feel
## V041109A     HHListing.9a. Respondent gender
## V043116      J1x. Summary: R party ID
## V043250      Y1x. Summary: Respondent age
```

This was a surprise to me



Create a New Dependent Variable

The difference in thermometer scores:

```
mydta1$th.bush.kerry <- mydta1$V043038 - mydta1$V043039
```

Clean up a bunch of variables & value labels

```
##Party  
mydta1$V043116 <- mydta1$V043116[, drop = TRUE]  
levels(mydta1$V043116) <- c("SD", "WD", "ID", "I", "IR", "WR", "  
  SR", "O")  
mydta1$V043116[ mydta1$V043116 %in% levels(mydta1$V043116)  
  [8] ] <- NA  
mydta1$V043116 <- mydta1$V043116[, drop = TRUE]  
table(mydta1$V043116)
```

SD	WD	ID	I	IR	WR	SR
203	179	210	118	138	154	193

Create a New Dependent Variable ...

```
## IDEO
mydta1$V045117 <- mydta1$V045117[ , drop = TRUE]
levels(mydta1$V045117) <- c("EL", "L", "SL", "M", "SC", "C", "EC"
)
table(mydta1$V045117)
```

EL	L	SL	M	SC	C	EC
20	103	125	279	143	166	31

```
## Gender
levels(mydta1$V041109A) <- c("M", "F")
## Gay Marriage
levels(mydta1$V043210)
```

Create a New Dependent Variable ...

```
[1] "1. Should be allowed"
      "3. Should not be
      allowed"
[3] "5. Should not be allowed to marry but should be allowed
      " "VOL"
[5] "8. Don't know"
      "9. Refused"
```

```
mydta1$V043210[ mydta1$V043210 %in% levels(mydta1$V043210)
  [4:10] ] <-NA
mydta1$V043210 <- mydta1$V043210[, drop = TRUE]
levels(mydta1$V043210) <- c("Allow", "No", "Med")
## Economy
mydta1$V043213 <- mydta1$V043213[ , drop = TRUE]
l <- levels(mydta1$V043213)
econnew <- factor(mydta1$V043213, levels=c(l[2], l[3], l[1]),
  labels=c("Worse", "Same", "Better"))
table(mydta1$V043213, econnew)
```

Create a New Dependent Variable ...

	econnew		
	Worse	Same	Better
1. Better	0	0	190
3. Worse	668	0	0
5. The same	0	343	0

```
mydta1$V043213 <- econnew
rm(econnew)
##Flag
mydta1$V045145X <- mydta1$V045145X[, drop = TRUE]
```


What Should a Regression Table Look Like?

- It needs
 - coefficients
 - standard errors (or t-values, possibly)
 - model diagnostics like N and R^2 and so forth.
- I want it to be easy to generate nice looking tables automatically to make it easy to prepare presentations and class notes.

R Functions/Packages for Producing Regression Output

- `outreg`: An R function I prepared 2006, now in "rockchalk" package
- `memisc`: works well, output not quite "presentation ready"
- `xtable`: incomplete output, but latex or HTML works
- `apsrtable`: very similar to `outreg`
- `Hmisc` "latex" function
- `texreg`: a new regression table making framework

Are You A LaTeX User?

- Some scientists say it is required, same with math. Many of the smart people I know prepare documents in \LaTeX .
- if using MS Word or similar makes you feel like you are “finger painting” to format material, \LaTeX may be the right thing for you.
- My \LaTeX notes page: <http://pj.freefaculty.org/latex>
- My \LaTeX lecture notes and example documents are linked on that site.

Maybe we should stop and talk that over. Maybe I should show you my [LaTeX Overview](#)

Automatic Tables are Easiest for LaTeX users

- \LaTeX is structured document “markup”, and most table-making packages in R cater to \LaTeX users, rather than other folks.
- However, for the MS Word (Libre Office) addicted, all is not lost. Several packages now can export to HTML, which Word is often able to import gracefully.
- I'm preparing these lectures using Sweave, which means that R runs and inserts the tables in automatically, I literally never touch them.

Consider making 50 sets of tables, one for each student

- In my regression class, I create random data sets that are individualized to the students, and then I need to show them what their results ought to be.
- I could write 50 MS Word documents, one for each student, or
- I DID make a \LaTeX based script that automated the production of one report per student. Please inspect the result:
<http://pj.freefaculty.org/stat/ps706/pj-test2>
- Please examine one or two of those files and then tell me it would be easier to use MS Word...

Here's what To Do

Load the rockchalk package, in which the outreg function resides.

```
library(rockchalk)
mod1age <- lm(th.bush.kerry~V043250, data=
  mydta1)
outreg(mod1age, tight=F, modelLabels=c("Age
  as Predictor"))
```

Produces this LaTeX Markup

```

\begin{tabular}{*{3}{l}}
\hline
& \multicolumn{2}{c}{Age as  
Predictor} \\
& Estimate & (S.E.) \\
\hline
\hline
(Intercept) & -6.841 & (4.596) \\
V043250 & 0.184* & (0.092) \\
\hline
N & 1191 & \\
RMSE & 53.885 & \\
$R^2$ & 0.003 & \\
\hline
\hline

```

Which LaTeX Renders as

	Age as Predictor	
	Estimate	(S.E.)
(Intercept)	-6.841	(4.596)
V043250	0.184*	(0.092)
N	1191	
RMSE	53.885	
R^2	0.003	

* $p \leq 0.05$

Add Gender

```
## Run a new regression  
mod2age <- lm(th.bush.kerry~V043250 +  
  V041109A, data=mydata1)  
## Put 2 regressions in same table  
outreg(list(mod1age, mod2age), tight=T,  
  modelLabels=c("Age Only", "Age With Gender"  
  ))
```

My terminology: “tight” means coefficients and standard errors vertically aligned

Output To LaTeX

	Age Only Estimate (S.E.)	Age With Gender Estimate (S.E.)
(Intercept)	-6.841 (4.596)	-3.085 (4.831)
V043250	0.184* (0.092)	0.191* (0.092)
V041109AF	.	-7.713* (3.123)
N	1191	1191
RMSE	53.885	53.77
R^2	0.003	0.008
adj R^2	0.003	0.007

* $p \leq 0.05$

Recent updates to outreg

- outreg was the first useful R function I created, I was distributing it (sans packaging) since 2006.
- The rockchalk package includes it now, I've made some "user convenience" changes.
 - easier customization of model "header" labels and variable names
 - easier to customize the selection of "goodness of fit" indicators in the bottom of the table
 - I'm not an alpha 0.05 insister anymore, you can choose 1 or more alpha levels (with stars!)

outreg can create html file output

- This is a brand new feature in outreg 1.8 (June, 2013)
 - `outreg2HTML()` receives outreg results and converts into Web markup.
 - Word 2010 will not “paste special” the HTML markup I generate, but it can “Insert -> File” and it absorbs the HTML markup in a reasonable way. You can finger paint to customize.
 - Not as nice looking or as automatic as \LaTeX , but I would use it if somebody made me use MS Word.

Lets put that to the test

```
or1 <- outreg(list(mod1age, mod2age), tight=T,
  modelLabels=c("Age Only", "Age With Gender
  "))
```

```
\begin{tabular}{*{3}{l}}
\hline
& Age Only & Age With Gender
& \\
& Estimate & Estimate \\
& (S.E.) & (S.E.) \\
\hline
\hline
(Intercept) & -6.841 & -3.085 \\
& (4.596) & (4.831) \\
V043250 & 0.184* & 0.191* \\
& (0.092) & (0.092)
```

The xtable package is as old as R itself

- In the old old days, xtable was “THE” one to use for \LaTeX output
- I still use it to create quick tables of some summary output, particularly output from `rockchalk::summarizeNumeric`. (There’s an example in each file in `pj-test2` mentioned above).

xtable is Nice Too

This code:

```
library(xtable)
tabout1 <- xtable(mod2age)
print(tabout1, type="latex")
```

Generates A LaTeX Table

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.0850	4.8314	-0.64	0.5232
V043250	0.1913	0.0917	2.09	0.0371
V041109AF	-7.7134	3.1232	-2.47	0.0137

Maybe I'm missing something, but this is not really a presentable, finished table.

xtable: HTML Markup Output

This code:

```
print(tabout1, type="HTML")
```

Produces HTML Markup that another program can Absorb

```
<!-- html table generated in R 3.0.1 by xtable 1.7-1 package
-->
<!-- Tue May 21 18:37:48 2013 -->
<TABLE border=1>
<TR> <TH> </TH> <TH> Estimate </TH> <TH> Std. Error </TH>
  <TH> t value </TH> <TH> Pr(> |t|) </TH> </TR>
<TR> <TD align="right"> (Intercept) </TD> <TD align="right"
  "> -3.0850 </TD> <TD align="right"> 4.8314 </TD> <TD
  align="right"> -0.64 </TD> <TD align="right"> 0.5232 <
  /TD> </TR>
<TR> <TD align="right"> V043250 </TD> <TD align="right"> 0
  .1913 </TD> <TD align="right"> 0.0917 </TD> <TD align=
  "right"> 2.09 </TD> <TD align="right"> 0.0371 </TD> </
  TR>
<TR> <TD align="right"> V041109AF </TD> <TD align="right">
  -7.7134 </TD> <TD align="right"> 3.1232 </TD> <TD
  align="right"> -2.47 </TD> <TD align="right"> 0.0137 <
  /TD> </TR>
</TABLE>
```

And a nice LaTeX converter "toLatex"

- `memisc` (by Martin Elf) is a general purpose social science package that has many excellent features
- It is the **ONLY** program I use to make presentable "Cross Tabulation Tables"
 - It has excellent support for SPSS user therapy and post-addiction recovery.
 - So far as I know, `memisc` was the first R package to offer the special features that other packages are now emulating.
 - Easy ability to incorporate results from several regression models in a single table
 - Easy ability to export that result to a presentable \LaTeX format.

memisc: "mtable" Feature was a Major Breakthrough

```
library(memisc)
mod3age <- lm(th.bush.kerry~V043250 * V041109A, data=mydtal
)
mtable(mod1age , mod2age , mod3age)
```

Calls :

```
mod1age: lm(formula = th.bush.kerry ~ V043250, data = mydtal
)
mod2age: lm(formula = th.bush.kerry ~ V043250 + V041109A,
data = mydtal)
mod3age: lm(formula = th.bush.kerry ~ V043250 * V041109A,
data = mydtal)
```

	mod1age	mod2age	mod3age
(Intercept)	-6.841 (4.596)	-3.085 (4.831)	-6.536 (6.692)
V043250	0.184* (0.092)	0.191* (0.092)	0.265 (0.135)

memisc: "mtable" Feature was a Major Breakthrough ...

V041109A: F/M		-7.713*	-1.268
		(3.123)	(9.194)
V043250 x V041109A: F/M			-0.137
			(0.184)
R^2	0.003	0.008	0.009
adj. R^2	0.003	0.007	0.006
sigma	53.885	53.770	53.780
F	4.028	5.072	3.565
p	0.045	0.006	0.014
Log-likelihood	-6437.301	-6434.252	-6433.973
Deviance	3452404.627	3434769.503	3433162.607
AIC	12880.603	12876.504	12877.946
BIC	12895.850	12896.834	12903.359
N	1191	1191	1191

memisc also offers a nice LaTeX converter "toLatex"

```
toLatex ( mtable ( mod1age , mod2age , mod3age ) )
```

And the Result

	mod1age	mod2age	mod3age
(Intercept)	-6.841 (4.596)	-3.085 (4.831)	-6.841 (6.841)
V043250	0.184* (0.092)	0.191* (0.092)	0.184* (0.092)
V041109A: F/M		-7.713* (3.123)	-1.111 (9.111)
V043250 × V041109A: F/M			-0.111 (0.111)
R-squared	0.003	0.008	0.003
adj. R-squared	0.003	0.007	0.003
sigma	53.885	53.770	53.885
F	4.028	5.072	3.028
p	0.045	0.006	0.045
Log-likelihood	-6437.301	-6434.252	-6433.301
Deviance	3452404.627	3434769.503	3433162.627
AIC	12880.603	12876.504	12877.603
BIC	12895.850	12896.834	12903.850
N	1191	1191	1191

And the Result ...

Critique

- “Too Much” information at the bottom of the plot. I’d rather have the default be minimal and allow users to ask for more.
- No symbol key to indicate what * ** mean
- Variable Labels difficult to understand

rms has offered \LaTeX for 20 years

- The package now known as “rms” (Regression Modeling Strategies, by Frank Harrell) has been offering latex as a generic function for 20 years, even before R existed (because S did exist).
- Has many latex functions customized to the regression fitting routines in rms

apsrtable

- Prepared by a then-graduate student at my Alma Mater Washington University in St. Louis.
- Perhaps the title is unfortunate because it suggests that it is intended for the American Political Science Review, but that is not its only target audience.

texreg

- The most recent entry in the regression table sweepstakes, by Philip Leifeld
- It has specialized converters designed for many of the popular kinds of regression model, including many in packages I've never heard of.
 - At first I had admiration and awe at the dedication to write separate “back end” functions for all of the different kinds of regression
 - Later, I became angry that regression package writers do not standardize their object structures to be compatible with `glm()` and `lm()` in R base, necessitating a nearly heroic effort to translate each one, one by one.
 - After that, I decided it is crazy to cater to inconsistent package writers and tried to generalize `outreg` to accommodate that diversity without writing 100 separate little functions for 100 separate little regression packages.

Example texreg usage

```
library(texreg)
to1 <- texreg(list(mod1age, mod2age),
  return.string = TRUE, use.packages = FALSE
  , booktabs = TRUE, dcolumn = TRUE)
```

- One criticism: user interface to texreg almost unbelievably filled up with complicated looking options. Perhaps not for beginners in R or \LaTeX .

Example texreg output

	Model 1	Model 2
(Intercept)	-6.84 (4.60)	-3.09 (4.83)
V043250	0.18* (0.09)	0.19* (0.09)
V041109AF		-7.71* (3.12)
R ²	0.00	0.01
Adj. R ²	0.00	0.01
Num. obs.	1191	1191

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table: Statistical models

I wondered what that LaTeX Markup would look like

```
cat(to1)
```

```
\begin{table}
\begin{center}
\begin{tabular}{| D{.}{.}{4.3}@{} D{.}{.}{4.3}@{} }
\toprule
& \multicolumn{1}{c}{Model 1} & \multicolumn{1}{c}{
Model 2} \\
\midrule
(Intercept) & -6.84 & -3.09 & \\
& (4.60) & (4.83) & \\
V043250 & 0.18^{*} & 0.19^{*} & \\
& (0.09) & (0.09) & \\
V041109AF & & -7.71^{*} & \\
& & (3.12) & \\
\midrule
R$^2$ & 0.00 & 0.01 & \\
Adj. R$^2$ & 0.00 & 0.01 & \\
Num. obs. & 1191 & 1191 & \\
\bottomrule
\multicolumn{3}{l}{\scriptsize ***$p<0.001$}
,

```

One Extra Large Attraction for texreg

- The author, unlike others, has exerted himself to learn what CSS formatting and HTML magic MS Word can tolerate, and the function `htmlreg()` creates an output file that Word can import and it will look almost as good as in \LaTeX .
- On my system (where Word runs in an emulation layer for Linux), the resulting table did look nice, but it would tolerate no finger painting by me. I was not able to adjust anything. But that's probably more of a reflection on me than on `texreg`.