stargazer:
beautiful LaTeX and ASCII tables from R statistical output

Marek Hlavac*
Harvard University

September 6, 2013

1 Introduction

stargazer is an R package that creates LaTeX code and ASCII text for well-formatted regression tables, with multiple models side-by-side, as well as for summary statistics tables. It can also output the content of data frames directly into LaTeX.

2 Why Should I Use *stargazer*?

Compared to available alternatives, *stargazer* excels in at least three respects: its ease of use, the large number of models it supports, and its beautiful aesthetics. These advantages have made it the R-to-LaTeX package of choice for many satisfied users at research and teaching institutions around the world.

2.1 Ease of Use

stargazer was designed with the user's comfort in mind. The learning curve is very mild, and all arguments are very intuitive, so that even a beginning user of R or LaTeX can quickly become familiar with the package's many capabilities. The package is intelligent, and tries to minimize the amount of effort the user has to put into adjusting argument values.

If *stargazer* is given a set of regression model objects, for instance, the package will create a side-by-side regression table. By contrast, if the user feeds it a data frame, *stargazer* will know that the user is most likely looking for a summary statistics table or – if the *summary* argument is set to FALSE – wants to output the content of the data frame.

*Harvard University, Political Economy and Government; hlavac@fas.harvard.edu

A quick reproducible example shows just how easy *stargazer* is to use. You can install *stargazer* from CRAN in the usual way:

```
install.packages("stargazer")
library(stargazer)
```

To create a summary statistics table from the *'attitude'* data frame (which should be available with your default installation of R), simply run the following:

```
stargazer(attitude)
```

Table 1:

Statistic	N	Mean	St. Dev.	Min	Max
rating	30	64.633	12.173	40	85
complaints	30	66.600	13.315	37	90
privileges	30	53.133	12.235	30	83
learning	30	56.367	11.737	34	75
raises	30	64.633	10.397	43	88
critical	30	74.767	9.895	49	92
advance	30	42.933	10.289	25	72
high.rating	30	0.333	0.479	0	1

To output the contents of the first four rows of some data frame, specify the part of the data frame you would like to see, and set the *summary* option to FALSE:

```
stargazer(attitude[1:4,], summary=FALSE)
```

Table 2:

rating	complaints	privileges	learning	raises	critical	advance	high.rating
43	51	30	39	61	92	45	FALSE
63	64	51	54	63	73	47	FALSE
71	70	68	69	76	86	48	TRUE
61	63	45	47	54	84	35	FALSE

Now, let us try to create a simple regression table with three side-by-side models – two Ordinary Least Squares (OLS) and one probit regression model – using the *lm()* and *glm()* functions. We can set the *align* argument to TRUE, so that coefficients in each column are aligned along the decimal point. *Table 3* shows the result.

```
##  2 OLS models
linear.1 <- lm(rating ~ complaints + privileges + learning + raises + critical,
data=attitude)
linear.2 <- lm(rating ~ complaints + privileges + learning, data=attitude)
## create an indicator dependent variable, and run a probit model
attitude$high.rating <- (attitude$rating > 70)
probit.model <- glm(high.rating ~ learning + critical + advance, data=attitude,
family = binomial(link = "probit"))

stargazer(linear.1, linear.2, probit.model, title="Results", align=TRUE)
```

Table 3: Results

	Dependent variable:		
	rating		high.rating
	OLS		*probit*
	(1)	(2)	(3)
complaints	0.692***	0.682***	
	(0.149)	(0.129)	
privileges	−0.104	−0.103	
	(0.135)	(0.129)	
learning	0.249	0.238*	0.164***
	(0.160)	(0.139)	(0.053)
raises	−0.033		
	(0.202)		
critical	0.015		−0.001
	(0.147)		(0.044)
advance			−0.062
			(0.042)
Constant	11.011	11.258	−7.476**
	(11.704)	(7.318)	(3.570)
Observations	30	30	30
R^2	0.715	0.715	
Adjusted R^2	0.656	0.682	
Log likelihood			−9.087
Akaike Inf. Crit.			26.175
Residual Std. Error	7.139($df = 24$)	6.863($df = 26$)	
F statistic	12.063***($df = 5; 24$)	21.743***($df = 3; 26$)	
Note:			*p<0.1; **p<0.05; ***p<0.01

In *Table 4*, we go a little bit further, and make some formatting and labeling adjustments. In particular, we remove all empty lines from the table (using *no.space*), and use *omit.stat* to leave out several statistics – namely, the log-likelihood ("*LL*"), residual standard error ("*ser*") and the F-statistic ("*f*"). Additionally, we label each of the dependent and independent variables with an easy-to-understand name. To do so, we use the *dep.var.labels* and *covariate.labels* arguments. The result is a complex, publication-quality LaTeX table. The relevant command call looks like this:

```
stargazer(linear.1, linear.2, probit.model, title="Regression Results",
align=TRUE, dep.var.labels=c("Overall Rating","High Rating"),
covariate.labels=c("Handling of Complaints","No Special Privileges",
"Opportunity to Learn","Performance-Based Raises","Too Critical","Advancement"),
omit.stat=c("LL","ser","f"), no.space=TRUE)
```

Table 4: Regression Results

	Dependent variable:		
	Overall Rating		High Rating
	OLS		*probit*
	(1)	(2)	(3)
Handling of Complaints	0.692***	0.682***	
	(0.149)	(0.129)	
No Special Privileges	−0.104	−0.103	
	(0.135)	(0.129)	
Opportunity to Learn	0.249	0.238*	0.164***
	(0.160)	(0.139)	(0.053)
Performance-Based Raises	−0.033		
	(0.202)		
Too Critical	0.015		−0.001
	(0.147)		(0.044)
Advancement			−0.062
			(0.042)
Constant	11.011	11.258	−7.476**
	(11.704)	(7.318)	(3.570)
Observations	30	30	30
R^2	0.715	0.715	
Adjusted R^2	0.656	0.682	
Akaike Inf. Crit.			26.175
Note:			*p<0.1; **p<0.05; ***p<0.01

In *Table 5*, we limit ourselves to the two linear models, and report 90 percent confidence intervals (using *ci* and *ci.level*) instead of standard errors. In addition, we report the coefficients and confidence intervals on the same row (using *single.row*).

```
stargazer(linear.1, linear.2, title="Regression Results",
dep.var.labels=c("Overall Rating","High Rating"),
covariate.labels=c("Handling of Complaints","No Special Privileges",
"Opportunity to Learn","Performance-Based Raises","Too Critical","Advancement"),
omit.stat=c("LL","ser","f"), ci=TRUE, ci.level=0.90, single.row=TRUE)
```

Table 5: Regression Results

	Dependent variable:	
	Overall Rating	
	(1)	(2)
Handling of Complaints	0.692*** (0.447, 0.937)	0.682*** (0.470, 0.894)
No Special Privileges	−0.104 (−0.325, 0.118)	−0.103 (−0.316, 0.109)
Opportunity to Learn	0.249 (−0.013, 0.512)	0.238* (0.009, 0.467)
Performance-Based Raises	−0.033 (−0.366, 0.299)	
Too Critical	0.015 (−0.227, 0.258)	
Advancement	11.011 (−8.240, 30.262)	11.258 (−0.779, 23.296)
Observations	30	30
R^2	0.715	0.715
Adjusted R^2	0.656	0.682
Note:		*p<0.1; **p<0.05; ***p<0.01

To produce ASCII text output, rather than LaTeX code, we simply set the argument *type* to *"text"*:

```
stargazer(linear.1, linear.2, type="text", title="Regression Results",
dep.var.labels=c("Overall Rating","High Rating"),
covariate.labels=c("Handling of Complaints","No Special Privileges",
"Opportunity to Learn","Performance-Based Raises","Too Critical","Advancement"),
omit.stat=c("LL","ser","f"), ci=TRUE, ci.level=0.90, single.row=TRUE)
```

```
Regression Results
================================================================================
                                        Dependent variable:
                                ---------------------------------------------
                                             Overall Rating
                                       (1)                     (2)
--------------------------------------------------------------------------------
Handling of Complaints          0.692*** (0.447, 0.937) 0.682*** (0.470, 0.894)
No Special Privileges           -0.104 (-0.325, 0.118)  -0.103 (-0.316, 0.109)
Opportunity to Learn             0.249 (-0.013, 0.512)   0.238* (0.009, 0.467)
Performance-Based Raises        -0.033 (-0.366, 0.299)
Too Critical                     0.015 (-0.227, 0.258)
Advancement                     11.011 (-8.240, 30.262) 11.258 (-0.779, 23.296)
--------------------------------------------------------------------------------
Observations                              30                      30
R2                                       0.715                   0.715
Adjusted R2                              0.656                   0.682
================================================================================
Note:                                              *p<0.1; **p<0.05; ***p<0.01
```

Let us now change the order of the explanatory variables using the *order* argument, and remove the covariate labels. In particular, we would like *learning* and *privileges* to come before all the other covariates. In addition, of the summary statistics reported, let us keep only the number of observations (using the argument *keep.stat*). Instead of reporting ASCII text, we'll go back to producing LaTeX tables by returning the type argument to its default value of "latex". *Table 6* is our result.

```
stargazer(linear.1, linear.2, title="Regression Results",
dep.var.labels=c("Overall Rating","High Rating"),
order=c("learning", "privileges"),
keep.stat="n", ci=TRUE, ci.level=0.90, single.row=TRUE)
```

Table 6: Regression Results

	Dependent variable:	
	Overall Rating	
	(1)	(2)
learning	0.692*** (0.447, 0.937)	0.682*** (0.470, 0.894)
privileges	−0.104 (−0.325, 0.118)	−0.103 (−0.316, 0.109)
complaints	0.249 (−0.013, 0.512)	0.238* (0.009, 0.467)
raises	−0.033 (−0.366, 0.299)	
critical	0.015 (−0.227, 0.258)	
Constant	11.011 (−8.240, 30.262)	11.258 (−0.779, 23.296)
Observations	30	30
Note:		*p<0.1; **p<0.05; ***p<0.01

2.2 Including Custom Standard Errors

Instead of reporting the default standard errors, users can choose to include custom vectors. Let us take a look at a brief example, adopted – with permission – from Slawa Rokicki's excellent R for Public Health blog. In this example, we will consider the sales of ice cream. More specifically, we are going to analyze how they might be affected by the temperature outside. First, we will generate a data set that will, for 500 cities, include values of the following variables: temperature (variable *temp*); sales per 100,000 people (*sales*); and the proportion of the city's population that is female (*female*).

```
set.seed(5)
temp <- rnorm(500, mean = 80, sd = 12)
sales <- 2 + temp * 3

for (i in 1:length(sales)) {
 if (temp[i]<75 | temp[i]>95) sales[i] <- sales[i] + rnorm(1, 0, 25)|
 else sales[i] <- sales[i] + rnorm(1, 0, 8)
}

female <- rnorm(500, mean = 0.5, sd = 0.01)
icecream <- as.data.frame(cbind(temp, sales, female))
```

Now, let us run a simple Ordinary Least Squares (OLS) regression model, and use the *sandwich* package to obtain heteroskedasticity-robust standard errors:

```
reg.model <- lm(sales ~ temp + female, data = icecream)

library(sandwich)
cov <- vcovHC(reg.model, type = "HC")
robust.se <- sqrt(diag(cov))
```

We can now use stargazer to create a regression table with the default and heteroskedasticity-robust standard errors in two columns, side by side:

```
stargazer(reg.model, reg.model, se=list(NULL, robust.se))
```

Table 7:

	Dependent variable:	
	sales	
	(1)	(2)
temp	2.972***	2.972***
	(0.067)	(0.084)
female	−37.819	−37.819
	(81.064)	(78.926)
Constant	23.916	23.916
	(41.187)	(40.408)
Observations	500	500
R^2	0.799	0.799
Adjusted R^2	0.798	0.798
Residual Std. Error (df = 497)	18.068	18.068
F Statistic (df = 2; 497)	984.916***	984.916***
Note:	*p<0.1; **p<0.05; ***p<0.01	

2.3 Many supported models

stargazer supports objects from the most widely used statistical functions and packages. In particular, the package supports model objects from *aftreg* (eha), *betareg* (betareg), *binaryChoice* (sampleSelection), *bj* (rms), *brglm* (brglm), *coeftest* (lmtest), *coxph* (survival), *coxreg* (eha), *clm* (ordinal), *clogit* (survival), *cph* (rms), *dynlm* (dynlm), *ergm* (ergm), *errorsarlm* (spdev), *gam* (mgcv), *gee* (gee), *glm* (stats), *Glm* (rms), *glmer* (lme4), *gls* (nlme), *Gls* (rms), *gmm* (gmm), *heckit* (sampleSelection), *hurdle* (pscl), *ivreg* (AER), *lagarlm* (spdep), *lm* (stats), *lmer* (lme4), *lmrob* (robustbase), *lrm* (rms), *maBina* (erer), *mclogit* (mclogit), *mlogit* (mlogit), *mlreg* (eha), *multinom* (nnet), *nlmer* (lme4), *ols* (rms), *phreg* (eha), *plm* (plm), *pmg* (plm), *polr* (MASS), *psm* (rms), *rem.dyad* (relevent), *rlm* (MASS), *rq* (quantreg), *Rq* (rms), *selection* (sampleSelection), *svyglm* (survey), *survreg* (survival), *tobit* (AER), *weibreg* (eha), *zeroinfl* (pscl), as well as from the implementation of these

in *zelig*. In addition, stargazer also supports the following zelig models: "*relogit*", "*cloglog.net*", "*gamma.net*", "*probit.net*" and "*logit.net*".

The number of models and objects that *stargazer* can accommodate puts it ahead of most of the alternative R-to-LaTeX options. As the development of the package continues, this list will continue expanding to matching models, as well as new, user-made, or customized statistical models.

2.4 Beautiful aesthetics

stargazer is very pleasing to the eye, and allows the user to customize all variable labels, as well as the formatting of the resulting table. If you'd like to create tables that look like those from your discipline's leading journal, *stargazer* can help you with that as well. You can use the style argument to choose a template of your choice. Economics and management scholars can thus create tables that resemble those published in the *American Economic Review*, in the *Quarterly Journal of Economics*, or in *Administrative Science Quarterly*. Political scientists can avail themselves of templates based on the *American Political Science Review*, the *American Journal of Political Science*, and on *International Organization*. For sociologists and demographers, the *American Sociological Review*, the *American Sociological Review* and *Demography* are available.

3 Citing *stargazer* in Research Publications

If you use the *stargazer* package in your research publications, please remember to include the following citation:

```
Hlavac, Marek (2013). stargazer: LaTeX code and ASCII text for well-formatted re-
gression and summary statistics tables. R package version 4.5.
http://CRAN.R-project.org/package=stargazer
```

Note: An early version of this document was adapted from my guest blog post on Tal Galili's excellent R-statistics blog.