stargazer : LaTeX code for well-formatted regression and summary statistics tables (R package)

Marek Hlavac

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

Overview

stargazer produces LaTeX code for well-formatted tables that hold regression analysis results from several models side-by-side, as well as summary statistics. It supports model objects from *lm*, *glm*, *svyglm*, *gee*, *gam*, *polr*, *survreg*, *coxph*, as well as from the implementation of these in *zelig*. It also supports the following *zelig* models for social network analysis: "cloglog.net", "gamma.net", and "logit.net".

How can I install stargazer?

stargazer is available from the Comprehensive R Archive Network (CRAN), along with its documentation in Adobe PDF format:
http://cran.r-project.org/web/packages/stargazer/index.html

Please make sure your version of R is up-to-date. You need to be running **R version 2.14 or newer** in order to install **stargazer**. You can download the latest version of R from: *http://www.r-project.org/*

To install this software, type the following command into the R prompt:
> **install.packages("stargazer")**

You will then be able to load the package in your R programs using:
> **library(stargazer)**

Supported journal styles

In addition to its drop-dead gorgeous default style, **stargazer** can create tables similar to those published in the following top economics, political science, management, public policy, sociology and demography journals:

American Economic Review (AER)
American Journal of Political Science (AJPS)
American Journal of Sociology (AJS)
Administrative Science Quarterly (ASQ)
American Sociological Review (ASR)
American Political Science Review (APSR)
Demography
International Organization (IO)
Journal of Policy Analysis and Management (JPAM)
Quarterly Journal of Economics (QJE)

Usage

All of the package's functionality is accessible through its single function **stargazer()**.

stargazer(..., title = "", style = "default", covariate.labels = NULL, dep.var.labels = NULL, decimal.mark = NULL, digit.separate = NULL, digit.separator = NULL, digits = NULL, digits.extra = NULL, initial.zero = NULL, intercept.top = NULL, model.names = NULL, model.numbers = NULL, notes = NULL, notes.align = NULL, notes.label = NULL, omit = NULL, omit.labels = NULL, omit.yes.no = c("Yes", "No"), ord.intercepts = FALSE, star.char = NULL, star.cutoffs = NULL, nobs = TRUE, mean.sd = TRUE, min.max = TRUE, median = FALSE, iqr = FALSE)

Arguments

... : one or more model objects (for regression analysis tables) or data frames (for summary statistics).
title : character string vector with titles for the tables.
style : character string that specifies what style, typically designed to resemble an existing academic journal, should be used in producing the tables.
covariate.labels: a character string vector of labels for covariates in regression tables.
dep.var.labels : a character string vector of labels for the dependent variables in regression tables.
decimal.mark : character string that will serve as the decimal mark. For instance, the string "," will represent decimal commas, while "." means tables will use decimal points.
digit.separate : a numerical vector that indicates where digit separators should be placed.
digit.separator : character string that will serve as the digit (e.g., thousands) separator.
digits : integer that indicates how many decimal places should be used.
digits.extra : integer indicating the maximum number of additional decimal places to be used if a number, rounded to digits decimal places, is equal to zero.
initial.zero : a logical value indicating whether an initial zero should be printed before the decimal mark if a number is between 0 and 1.
intercept.top : a logical value indicating whether the intercept (or constant) coefficients should be on top, rather than on the bottom, of the table.
model.names : a logical value indicating whether model names should be included in the table.
model.numbers : a logical value indicating whether models should be numbered. No number is used whenever a regression table includes only one model.
notes : a character string vector containing notes to be included below the table.
notes.align : a character string that specifies how notes should be aligned under the table.
notes.label : a character string containing a label for the notes section of the table.
omit : a vector of regular expressions that specifies which of the explanatory variables should be omitted from presentation in the table. This argument might be used, for instance, to exclude fixed effects dummies from being presented. The default value of NULL means that no variables will be excluded.
omit.labels : a character string vector of labels that correspond to each of the regular expressions in omit, and that will be used in a sub-table that indicates whether variables have been omitted from a given model.
omit.yes.no : a character string vector of length 2 that contains the 'yes' and 'no' strings to indicate whether, in any specific model, variables were omitted from the table, as specified by "omit".
ord.intercepts : a logical value indicating whether intercepts for models with ordered dependent variables (such as ordered probit, or ordered logit) are included in the table.
star.char : character string to be used as the 'star' to denote statistical significance.
star.cutoffs : a numeric vector with a length of at least 1 and at most 3 that indicates the statistical significance cutoffs for one, two and three 'stars,' respectively. For elements with NA values, the corresponding 'star' will not be used.
nobs : a logical value that toggles whether the number of observations (N) for each variable is shown in summary statistics tables.
mean.sd : a logical value that toggles whether variable means and standard deviations are shown in summary statistics tables.
min.max : a logical value that toggles whether variable minima and maxima are shown in summary statistics tables.
median : a logical value that toggles whether variable medians are shown in summary statistics tables.
iqr: a logical value that toggles whether the 25th and 75th percentiles for each variable are shown in summary statistics tables. ('iqr' stands for interquartile range.)

Examples

```
## create summary statistics table based for 'attitude' data frame
stargazer(attitude)
##estimate and report results of 3 OLS models
m1 <- lm(rating  complaints + privileges + learning + raises + critical, data=attitude)
m2 <- lm(rating  complaints + privileges + learning, data=attitude)
m3 <- lm(rating  learning + critical + advance, data=attitude)
stargazer(m1, m2, m3)
```

Prepared for the Second Annual Government Department Poster Session, Harvard University, 16 October 2012. I am grateful to Erin Baggott, Volha Charnych, John Coglianese, Stefan Dimitriadis, Cheng Gao, Mallory James, Melissa Kline, Carl J. Liebersohn, Christopher Lucas, Stephen Pettigrew, Jeniffer Sheehy-Skeffington, Emily Stephen, Beth Truesdale and Daniel Yew Mao Lim for their assistance.