## **Model Summaries**

## Table of contents

1	output: print and save	3
2	fmt: round and format	5
3	estimate	6
4	statistic: SE, t, p, CI, etc.	8
5	vcov: robust SEs	12
6	stars	13
7	coef_omit	14
8	coef_rename	16
9	coef_map: order, omit, rename	20
10	gof_omit: goodness-of-fit and model characteristics	20
11	gof_map	21
12	shape: pivot, groups, panels, and stacks  12.1 Formula	22 23 27 29 29 30
13	align: column alignment	32
14	notes: footers	33

15	title: captions	33
16	add_rows	34
17	Math	34
18	exponentiate	36
19	ellipsis: Additional arguments	36
20	Custom appearance	38
21	Supported models	38
22	Rmarkdown, Quarto, Org-Mode         22.1 Rmarkdown          22.2 Quarto          22.3 Emacs Org-Mode	42 42 42 43
23	Global options	44
24	Case studies 24.1 Reference categories 24.2 Standardization 24.3 Subgroup estimation with nest_by 24.4 Statistics in separate columns instead of one over the other 24.5 Bootstrap 24.6 fixest: Fixed effects and instrumental variable regression 24.7 Multiple imputation	45 46 48 48 50 51 53
25	Table-making packages	54
26	FAQ 26.1 Stack Overflow is useful	55 56 56 56 57
	26.6 Escaped LaTeX characters	59 59

modelsummary includes a powerful set of utilities to customize the information displayed in your model summary tables. You can easily rename, reorder, subset or omit parameter estimates; choose the set of goodness-of-fit statistics to display; display various "robust" standard errors or confidence intervals; add titles, footnotes, or source notes; insert stars or custom characters

to indicate levels of statistical significance; or add rows with supplemental information about your models.

Before starting, we set modelsummary to use the tinytable package to draw table, and we set a few CSS properties to make tables more display in a more compact format on this website. This step is optional.

Now, we download data, fit some models, and summarize them in a table using the modelsummary() function:

```
library(modelsummary)
library(tinytable)

url <- 'https://vincentarelbundock.github.io/Rdatasets/csv/HistData/Guerry.csv'
dat <- read.csv(url)

models <- list(
   "OLS 1" = lm(Donations ~ Literacy + Clergy, data = dat),
   "Poisson" = glm(Donations ~ Literacy + Commerce, family = poisson, data = dat),
   "OLS 2" = lm(Crime_pers ~ Literacy + Clergy, data = dat)
)

modelsummary(models)</pre>
```

#### 1 output: print and save

The output argument determines the type of object returned by modelsummary and/or the file where this table should be written.

If you want to save a table directly to file, you can type:

```
modelsummary(models, output = "table.docx")
modelsummary(models, output = "table.html")
modelsummary(models, output = "table.tex")
modelsummary(models, output = "table.md")
modelsummary(models, output = "table.txt")
modelsummary(models, output = "table.png")
```

If you want a raw HTML, LaTeX, or Markdown table, you can type:

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16259.384
	(2078.276)	(0.006)	(2611.140)
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148
	(25.735)		(32.334)
Commerce		0.011	
		(0.000)	
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

```
modelsummary(models, output = "html")
modelsummary(models, output = "latex")
modelsummary(models, output = "markdown")
```

If you to customize the appearance of your table using external tools like gt, kableExtra, flextable, or huxtable, you can type:

```
modelsummary(models, output = "gt")
modelsummary(models, output = "kableExtra")
modelsummary(models, output = "flextable")
modelsummary(models, output = "huxtable")
```

Warning: When a file name is supplied to the output argument, the table is written immediately to file. If you want to customize your table by post-processing it with an external package, you need to choose a different output format and saving mechanism. Unfortunately, the approach differs from package to package:

- tinytable: set output="tinytable", post-process your table, and use the tinytable::save\_tt() function.
- gt: set output="gt", post-process your table, and use the gt::gtsave() function.
- kableExtra: set output to your destination format (e.g., "latex", "html", "markdown"), post-process your table, and use kableExtra::save\_kable() function.

#### 2 fmt: round and format

The fmt argument defines how numeric values are rounded and presented in the table. This argument accepts three types of input:

- Integer: Number of decimal digits
- User-supplied function: Accepts a numeric vector and returns a character vector of the same length.
- modelsummary function: fmt\_decimal(), fmt\_significant(), fmt\_sprintf(), fmt\_term(), fmt\_statistic, fmt\_identity()

#### Examples:

```
mod <- lm(mpg ~ hp + drat + qsec, data = mtcars)
# decimal digits
modelsummary(mod, fmt = 3)</pre>
```

```
# user-supplied function
modelsummary(mod, fmt = function(x) round(x, 2))

# p values with different number of digits
modelsummary(mod, fmt = fmt_decimal(1, 3), statistic = c("std.error", "p.value"))

# significant digits
modelsummary(mod, fmt = fmt_significant(3))

# sprintf(): decimal digits
modelsummary(mod, fmt = fmt_sprintf("%.5f"))

# sprintf(): scientific notation
modelsummary(mod, fmt = fmt_sprintf("%.5e"))

# statistic-specific formatting
modelsummary(mod, fmt = fmt_statistic(estimate = 4, conf.int = 1), statistic = "conf.int")

# term-specific formatting
modelsummary(mod, fmt = fmt_term(hp = 4, drat = 1, default = fmt_significant(2)))
modelsummary(mod, fmt = NULL)
```

In many languages the comma is used as a decimal mark instead of the period. modelsummary respects the global R OutDec option, so you can simply execute this command and your tables will be adjusted automatically:

```
options(OutDec=",")
```

#### 3 estimate

By default, modelsummary prints each coefficient estimate on its own row. You can customize this by changing the estimate argument. For example, this would produce a table of p values instead of coefficient estimates:

```
modelsummary(models, estimate = "p.value")
```

You can also use glue string, using curly braces to specify the statistics you want. For example, this displays the estimate next to a confidence interval:

	OLS 1	Poisson	OLS 2
Literacy	-39.1 [-112.8, 34.6]	0.0 [0.0, 0.0]	3.7 [-88.9, 96.3]
Clergy	15.3 [-35.9, 66.4]		77.1 [12.8, 141.5]
Commerce		$0.0 \ [0.0, \ 0.0]$	
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

```
modelsummary(
  models,
  fmt = 1,
  estimate = "{estimate} [{conf.low}, {conf.high}]",
  statistic = NULL,
  coef_omit = "Intercept")
```

Glue strings can also apply R functions to estimates. However, since modelsummary rounds numbers and transforms them to character by default, we must set fmt = NULL:

```
m <- glm(am ~ mpg, data = mtcars, family = binomial)
modelsummary(
    m,
    fmt = NULL,
    estimate = "{round(exp(estimate), 5)}",
    statistic = "{round(exp(estimate) * std.error, 3)}")</pre>
```

You can also use different estimates for different models by using a vector of strings:

	(1)
(Intercept)	0.00136
	0.003
mpg	1.35938
	0.156
Num.Obs.	32
AIC	33.7
BIC	36.6
Log.Lik.	-14.838
F	7.148
RMSE	0.39

```
"{estimate} ({std.error})"),
statistic = NULL,
coef_omit = "Intercept")
```

### 4 statistic: SE, t, p, CI, etc.

By default, modelsummary prints the coefficient's standard error in parentheses below the corresponding estimate. The value of this uncertainty statistic is determined by the statistic argument. The statistic argument accepts any of the column names produced by get\_estimates(model). For example:

```
modelsummary(models, statistic = 'std.error')
modelsummary(models, statistic = 'p.value')
modelsummary(models, statistic = 'statistic')
```

You can also display confidence intervals in brackets by setting statistic="conf.int":

Alternatively, you can supply a glue string to get more complicated results:

	OLS 1	Poisson	OLS 2
Literacy	-39.1	0.0***	3.7 (46.6)
Clergy	15.3		77.1 (32.3)
Commerce		0.0***	
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

	OLS 1	Poisson	OLS 2
(Intercept)	7948.7	8.2	16 259.4
	[2469.6,  13427.8]	[8.2, 8.3]	$[9375.5,\ 23\ 143.3]$
Literacy	-39.1	0.0	3.7
	[-136.8, 58.6]	[0.0, 0.0]	[-119.0, 126.4]
Clergy	15.3		77.1
	[-52.6, 83.1]		[-8.1, 162.4]
Commerce		0.0	
		[0.0,  0.0]	
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16 259.384
	$2078.276 \ (< 0.001)$	$0.006 \ (< 0.001)$	2611.140 (<0.001)
Literacy	-39.121	0.003	3.680
	$37.052\ (0.294)$	0.000 (< 0.001)	$46.552 \ (0.937)$
Clergy	15.257		77.148
	$25.735 \ (0.555)$		$32.334 \ (0.019)$
Commerce		0.011	
		0.000 (<0.001)	
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

You can also display several different uncertainty estimates below the coefficient estimates by using a vector. For example,

Setting statistic=NULL omits all statistics. This can often be useful if, for example, you want to display confidence intervals next to coefficients:

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16 259.384
	[3815.060,12082.275]	$[8.230,\ 8.252]$	[11065.933,21452.836]
	s.e. $= 2078.276$	s.e. $= 0.006$	s.e. $= 2611.140$
	t = 3.825	t = 1408.907	t = 6.227
	p = < 0.001	p = < 0.001	p = < 0.001
Literacy	-39.121	0.003	3.680
	$[-112.816,\ 34.574]$	[0.003,  0.003]	[-88.910, 96.270]
	s.e. $= 37.052$	s.e. $= 0.000$	s.e. $= 46.552$
	t = -1.056	t = 33.996	t = 0.079
	p = 0.294	p = < 0.001	p = 0.937
Clergy	15.257		77.148
	[-35.930,66.443]		$[12.837,\ 141.459]$
	s.e. $= 25.735$		s.e. $= 32.334$
	t = 0.593		t = 2.386
	p = 0.555		p = 0.019
Commerce		0.011	
		[0.011,  0.011]	
		s.e. $= 0.000$	
		t = 174.542	
		p = <0.001	

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667 [3815.060, 12082.275]	8.241 [8.230, 8.252]	16 259.384 [11 065.933, 21 452.836]
Literacy	-39.121 [ $-112.816$ , $34.574$ ]	$0.003 \ [0.003, \ 0.003]$	3.680 [-88.910, 96.270]
Clergy	15.257 [-35.930, 66.443]		77.148 [12.837, 141.459]
Commerce		0.011 [0.011, 0.011]	

#### 5 vcov: robust SEs

You can use clustered or robust uncertainty estimates by modifying the vcov parameter. This function accepts 5 different types of input. You can use a string or a vector of strings:

```
modelsummary(models, vcov = "robust")
modelsummary(models, vcov = c("classical", "HC3", "bootstrap"))
```

These variance-covariance matrices are calculated using the sandwich package. You can pass arguments to the sandwich functions directly from the modelsummary function. For instance, to change the number of bootstrap replicates and to specify a clustering variable we could call:

```
modelsummary(mod, vcov = "bootstrap", R = 1000, cluster = "country")
```

You can use a one-sided formula or list of one-sided formulas to use clustered standard errors:

```
modelsummary(models, vcov = ~Region)
```

You can specify a function that produces variance-covariance matrices:

```
library(sandwich)
modelsummary(models, vcov = vcovHC)
```

You can supply a list of functions of the same length as your model list:

```
modelsummary(models,
    vcov = list(vcov, vcovHC, vcovHC, vcovH))
```

You can supply a list of named variance-covariance matrices:

```
vcov_matrices <- lapply(models, vcovHC)
modelsummary(models, vcov = vcov_matrices)</pre>
```

You can supply a list of named vectors:

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667***	8.241***	16 259.384***
	(2078.276)	(0.006)	(2611.140)
Literacy	-39.121	0.003***	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148*
	(25.735)		(32.334)
Commerce		0.011***	
		(0.000)	

#### 6 stars

Some people like to add "stars" to their model summary tables to mark statistical significance. The stars argument can take three types of input:

- 1. NULL omits any stars or special marks (default)
- 2. TRUE uses these default values: + p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001
- 3. Named numeric vector for custom stars.

```
modelsummary(models)
modelsummary(models, stars = TRUE)
modelsummary(models, stars = c('+' = .1, '&' = .01))
```

Whenever stars is not NULL, modelsummary adds a note at the bottom of the table automatically. If you would like to add stars but *not* include a note at the bottom of the table, you can define the display of your estimate manually using a glue string, as described in the estimate argument section of the documentation. Whenever the {stars} string appears in the estimate or statistic arguments, modelsummary will assume that you want fine-grained control over your table, and will *not* include a note about stars.

If you want to create your own stars description, you can add custom notes with the **notes** argument.

	OLS 1	Poisson	OLS 2
Clergy	15.257		77.148
	(25.735)		(32.334)
Commerce		0.011	
		(0.000)	

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16 259.384
	(2078.276)	(0.006)	(2611.140)
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)

## 7 coef\_omit

An alternative mechanism to subset coefficients is to use the <code>coef\_omit</code> argument, which accepts a vector of integer or a regular expression. For example, we can omit the first and second coefficients as follows:

```
modelsummary(models, coef_omit = 1:2, gof_map = NA)
```

Negative indices determine which coefficients to *keep*:

```
modelsummary(models, coef_omit = c(-1, -2), gof_map = NA)
```

When coef\_omit is a string, it is fed to grepl(x,perl=TRUE) to detect the variable names which should be excluded from the table.

```
modelsummary(models, coef_omit = "Intercept|.*merce", gof_map = NA)
```

	OLS 1	Poisson	OLS 2
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148
	(25.735)		(32.334)

	OLS 1	Poisson	OLS 2
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
	OLS 1	Poisson	OLS 2
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148
	(25.735)		(32.334)

Since coef\_omit accepts regexes, you can do interesting things with it, such as specifying the list of variables that modelsummary should *keep* instead of omit. To do this, we use a negative lookahead. To keep only the coefficients *starting* with "Lit", we call:

```
modelsummary(models, coef_omit = "^(?!Lit)", gof_map = NA)
```

To keep all coefficients matching the "y" substring:

```
modelsummary(models, coef_omit = "^(?!.*y)", gof_map = NA)
```

To keep all coefficients matching one of two substrings:

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16259.384
	(2078.276)	(0.006)	(2611.140)
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148
	(25.735)		(32.334)

	(1)	(2)
(Intercept)	353.653	189.722
	(76.049)	(11.347)
Explanator	-57.545	-98.365
	(20.922)	(17.155)
Num.Obs.	32	32
R2	0.201	0.523
R2 Adj.	0.175	0.507
AIC	359.2	342.7
BIC	363.6	347.1
Log.Lik.	-176.588	-168.347
F	7.565	32.876
RMSE	60.31	46.61

#### 8 coef\_rename

modelsummary offers powerful and innovative mechanisms to rename, reorder, and subset coefficients and goodness-of-fit statistics.

You can rename coefficients using the coef\_rename argument. For example, if you have two models with different explanatory variables, but you want both variables to have the same name and appear on the same row, you can do:

If you provide a named character vector to coef\_rename, only exact matches of the *complete* original term name will be replaced.

For complex modifications, you can feed a function which returns a named vector to the coef\_rename argument. For example, modelsummary ships with a function called coef\_rename, which executes some common renaming tasks automatically. This example also uses the dvnames function to extract the name of the dependent variable in each model:

```
x <- list(
lm(mpg ~ factor(cyl) + drat + disp, data = mtcars),</pre>
```

	mpg	hp
(T. )		
(Intercept)	26.158	-86.788
	(6.537)	(79.395)
6	-4.547	46.485
	(1.731)	(21.027)
8	-4.580	121.892
	(2.952)	(35.853)
Drat	0.783	37.815
	(1.478)	(17.952)
Disp	-0.026	0.147
	(0.011)	(0.137)
Num.Obs.	32	32
R2	0.786	0.756
R2 Adj.	0.754	0.720
AIC	167.4	327.2
BIC	176.2	336.0
Log.Lik.	-77.719	-157.623
F	24.774	20.903
RMSE	2.74	33.34

```
lm(hp ~ factor(cyl) + drat + disp, data = mtcars)
)
modelsummary(dvnames(x), coef_rename = coef_rename)
```

Of course, you can also define your own custom functions. For instance, to rename a model with interacted variables (e.g., "drat:mpg"), you could define a custom rename\_explanator function:

```
y <- list(
  lm(hp ~ drat / mpg, mtcars),
  lm(hp ~ vs / mpg, mtcars)
)
rename_explanator <- function(old_names) {</pre>
```

	(1)	(2)
(Intercept)	91.206	189.722
	(72.344)	(11.205)
Explanator	68.331	-18.316
	(27.390)	(62.531)
Explanator:mpg	-2.558	-3.260
	(0.467)	(2.451)
Num.Obs.	32	32
R2	0.608	0.550
R2 Adj.	0.581	0.519
AIC	338.4	342.8
BIC	344.3	348.7
Log.Lik.	-165.218	-167.399
F	22.454	17.743
RMSE	42.27	45.25

```
new_names <- gsub("drat|vs", "Explanator", old_names)
setNames(new_names, old_names)
}
modelsummary(y, coef_rename = rename_explanator)</pre>
```

Beware of inadvertently replacing parts of other variable names! Making your regex pattern as specific as possible (e.g., by adding word boundaries) is likely a good idea. The custom rename function is also a good place to re-introduce the replacement of ":" with "×" if you are dealing with interaction terms — modelsummary makes this replacement for you only when the coef\_rename argument is not specified.

Another possibility is to assign variable labels to attributes in the data used to fit the model. Then, we can automatically rename them:

```
datlab <- mtcars
datlab$cyl <- factor(datlab$cyl)
attr(datlab$cyl, "label") <- "Cylinders"
attr(datlab$am, "label") <- "Transmission"
modlab <- lm(mpg ~ cyl + am, data = datlab)
modelsummary(modlab, coef_rename = TRUE)</pre>
```

	(1)
(Intercept)	24.802
	(1.323)
Cylinders [6]	-6.156
	(1.536)
Cylinders [8]	-10.068
	(1.452)
Transmission	2.560
	(1.298)
Num.Obs.	32
R2	0.765
R2 Adj.	0.740
AIC	168.4
BIC	175.7
Log.Lik.	-79.199
$\mathbf{F}$	30.402
RMSE	2.87

	OLS 1	Poisson	OLS 2
Literacy (%)	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Patents per capita		0.011	
		(0.000)	
Constant	7948.667	8.241	16259.384
	(2078.276)	(0.006)	(2611.140)
Num.Obs.	86	86	86
R2	0.020		0.065
R2 Adj.	-0.003		0.043
AIC	1740.8	274160.8	1780.0
BIC	1750.6	274168.2	1789.9
Log.Lik.	-866.392	-137077.401	-886.021
F	0.866	18294.559	2.903
RMSE	5740.99	5491.61	7212.97

## 9 coef\_map: order, omit, rename

The coef\_map argument is a named vector which allows users to rename, reorder, and subset coefficient estimates. Values of this vector correspond to the "clean" variable name. Names of this vector correspond to the "raw" variable name. The table will be sorted in the order in which terms are presented in coef\_map. Coefficients which are *not* included in coef\_map will be excluded from the table.

## 10 gof\_omit: goodness-of-fit and model characteristics

gof\_omit is a regular expression which will be fed to grepl(x,perl=TRUE) to detect the names of the statistics which should be excluded from the table.

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16 259.384
	(2078.276)	(0.006)	(2611.140)
Literacy	-39.121	0.003	3.680
	(37.052)	(0.000)	(46.552)
Clergy	15.257		77.148
	(25.735)		(32.334)
Commerce		0.011	
		(0.000)	
Num.Obs.	86	86	86
R2	0.020		0.065

```
modelsummary(models, gof_omit = 'DF|Deviance|R2|AIC|BIC')
```

#### 11 gof\_map

The gof\_map argument can be used to rename, re-order, subset, and format the statistics displayed in the bottom section of the table ("goodness-of-fit").

The first type of values allowed is a character vector with elements equal to column names in the data.frame produced by get\_gof(model):

```
modelsummary(models, gof_map = c("nobs", "r.squared"))
```

A more powerful mechanism is to supply a data.frame (or tibble) through the gof\_map argument. This data.frame must include 3 columns:

- 1. raw: a string with the name of a column produced by get\_gof(model).
- 2. clean: a string with the "clean" name of the statistic you want to appear in your final table.
- 3. fmt: a string which will be used to round/format the string in question (e.g., "%.3f"). This follows the same standards as the fmt argument in ?modelsummary.

You can see an example of a valid data frame by typing modelsummary::gof\_map. This is the default data.frame that modelsummary uses to subset and reorder goodness-of-fit statistics. As you can see, omit == TRUE for quite a number of statistics. You can include setting omit == FALSE:

```
gm <- modelsummary::gof_map
gm$omit <- FALSE
modelsummary(models, gof_map = gm)</pre>
```

The goodness-of-fit statistics will be printed in the table in the same order as in the gof\_map data.frame.

```
f <- function(x) format(round(x, 3), big.mark=",")
gm <- list(
  list("raw" = "nobs", "clean" = "N", "fmt" = f),
  list("raw" = "AIC", "clean" = "aic", "fmt" = f))
modelsummary(models, gof_map = gm)</pre>
```

Notice the subtle difference between coef\_map and gof\_map. On the one hand, coef\_map works as a "white list": any coefficient not explicitly entered will be omitted from the table. On the other, gof\_map works as a "black list": statistics need to be explicitly marked for omission.

Another convenient way to build a gof\_map argument is to use the tribble function from the tibble package. In this example, we insert special HTML code to display a superscript, so we use the escape=FALSE argument.

Warning: This tribble syntax may not work well when fmt is a function. In that case, please revert to the list syntax illustrated above.

## 12 shape: pivot, groups, panels, and stacks

This section requires version 1.3.1 of modelsummary. If this version is not available on CRAN yet, you can install the development version by following the instructions on the website.

The shape argument accepts:

	OLS 1	Poisson	OLS 2
(Intercept)	7948.667	8.241	16 259.384
Literacy	-39.121	0.003	3.680
Clergy	15.257		77.148
Commerce		0.011	
N	86	86	86
R <sup>2</sup>	0.02		0.07

	(1)	(2)
(Intercept)	30.099	10.790
	(1.634)	(5.078)
hp	-0.068	-0.052
	(0.010)	(0.009)
drat		4.698
		(1.192)

- 1. A formula which determines the structure of the table, and can display "grouped" coefficients together (e.g., multivariate outcome or mixed-effects models).
- 2. The strings "cbind", "rbind" or "rcollapse" to stack multiple tables on top of each other and present models in distinct "panels".

#### 12.1 Formula

The left side of the formula represents the rows and the right side represents the columns. The default formula is term + statistic ~ model:

```
m <- list(
    lm(mpg ~ hp, data = mtcars),
    lm(mpg ~ hp + drat, data = mtcars))

modelsummary(m, shape = term + statistic ~ model, gof_map = NA)</pre>
```

We can display statistics horizontally with:

	(1)		(2)			
	Est.	2.5 %	97.5 %	Est.	2.5 %	97.5 %
(Intercept)	30.099	26.762	33.436	10.790	0.405	21.175
hp	-0.068	-0.089	-0.048	-0.052	-0.071	-0.033
drat				4.698	2.261	7.135

	E	st.	2.5	5 %	97.	5 %
	(1)	(2)	(1)	(2)	(1)	(2)
(Intercept)	30.099	10.790	26.762	0.405	33.436	21.175
hp	-0.068	-0.052	-0.089	-0.071	-0.048	-0.033
drat		4.698		2.261		7.135

```
statistic = "conf.int",
gof_map = NA)
```

The order of terms in the formula determines the order of headers in the table.

shape does partial matching and will try to fill-in incomplete formulas:

```
modelsummary(m, shape = ~ statistic)
```

Some models like multinomial logit or GAMLSS produce "grouped" parameter estimates. To display these groups, we can include a group identifier in the shape formula. This group identifier must be one of the column names produced by get\_estimates(model). For example, in models produced by nnet::multinom, the group identifier is called "response":

```
library(nnet)

dat_multinom <- mtcars
dat_multinom$cyl <- sprintf("Cyl: %s", dat_multinom$cyl)

mod <- list(</pre>
```

```
nnet::multinom(cyl ~ mpg, data = dat_multinom, trace = FALSE),
nnet::multinom(cyl ~ mpg + drat, data = dat_multinom, trace = FALSE))
get_estimates(mod[[1]])
```

```
term estimate std.error conf.level
                                              conf.low
                                                         conf.high statistic
1 (Intercept) 47.252432 34.975171
                                       0.95 -21.2976435 115.8025065 1.351028
         mpg -2.205418 1.637963
                                       0.95 -5.4157653
                                                         1.0049299 -1.346440
3 (Intercept) 72.440246 37.175162
                                       0.95 -0.4217332 145.3022247 1.948619
         mpg -3.579991 1.774693
                                       0.95 -7.0583242 -0.1016573 -2.017246
 df.error
             p.value response s.value group
      Inf 0.17668650
                      Cyl: 6
                                  2.5
1
2
      Inf 0.17816078 Cyl: 6
                                  2.5
3
      Inf 0.05134088 Cyl: 8
                                  4.3
      Inf 0.04366989
                       Cyl: 8
                                  4.5
```

To summarize the results, we can type:

```
# modelsummary(mod, shape = term + response ~ statistic)
```

The terms of the **shape** formula above can of course be rearranged to reshape the table. For example:

```
modelsummary(mod, shape = model + term ~ response)
```

We can combine the term and group identifier columns by inserting an interaction colon: instead of the + in the formula:

```
library(marginaleffects)
mod <- glm(am ~ mpg + factor(cyl), family = binomial, data = mtcars)
mfx <- avg_slopes(mod)

modelsummary(mfx, shape = term + contrast ~ model)</pre>
```

```
modelsummary(mfx, shape = term : contrast ~ model)
```

		Cyl: 6	Cyl: 8
(1)	(Intercept)	47.252	72.440
		(34.975)	(37.175)
	mpg	-2.205	-3.580
		(1.638)	(1.775)
(2)	(Intercept)	89.573	117.971
		(86.884)	(87.998)
	mpg	-3.627	-4.838
		(3.869)	(3.915)
	drat	-3.210	-5.028
		(3.810)	(4.199)

		(1)
cyl	6 - 4	0.097
		(0.166)
	8 - 4	0.093
		(0.234)
mpg	$\mathrm{d}Y/\mathrm{d}X$	0.056
		(0.027)
Num.Obs.		32
AIC		37.4
BIC		43.3
Log.Lik.		-14.702
F		2.236
RMSE		0.39

	(1)
cyl 6 - 4	0.097
	(0.166)
cyl 8 - 4	0.093
	(0.234)
$\mathrm{mpg}\ \mathrm{dY/dX}$	0.056
	(0.027)
Num.Obs.	32
AIC	37.4
BIC	43.3
Log.Lik.	-14.702
F	2.236
RMSE	0.39

# 12.2 String ("rbind" or "rcollapse"): Panels of models in stacked regression tables

Note: The code in this section requires version 1.3.0 or the development version of modelsummary. See the website for installation instructions.

This section shows how to "stack/bind" multiple regression tables on top of one another, to display the results several models side-by-side and top-to-bottom. For example, imagine that we want to present 4 different models, half of which are estimated using a different outcome variable. When using modelsummary, we store models in a list. When using modelsummary with shape="rbind" or shape="rbind", we store models in a list of lists:

```
gm <- c("r.squared", "nobs", "rmse")

panels <- list(
    list(
    lm(mpg ~ 1, data = mtcars),
    lm(mpg ~ qsec, data = mtcars)
),
    list(
    lm(hp ~ 1, data = mtcars),
    lm(hp ~ qsec, data = mtcars)
)</pre>
```

	(1)	(2)
(Intercept)	20.091	-5.114
	(1.065)	(10.030)
qsec		1.412
		(0.559)
R2	0.000	0.175
Num.Obs.	32	32
RMSE	5.93	5.39
(Intercept)	146.688	631.704
	(12.120)	(88.700)
qsec		-27.174
		(4.946)
R2	0.000	0.502
Num.Obs.	32	32
RMSE	67.48	47.64

```
modelsummary(
  panels,
  shape = "rbind",
  gof_map = gm)
```

Like with modelsummary(), we can can name models and panels by naming elements of our nested list:

```
panels <- list(
   "Outcome: mpg" = list(
       "(I)" = lm(mpg ~ 1, data = mtcars),
       "(II)" = lm(mpg ~ qsec, data = mtcars)
),
   "Outcome: hp" = list(
       "(I)" = lm(hp ~ 1, data = mtcars),
       "(II)" = lm(hp ~ qsec, data = mtcars)
)
modelsummary(</pre>
```

	(I)	(II)
(Intercept)	20.091	-5.114
	(1.065)	(10.030)
qsec		1.412
		(0.559)
R2	0.000	0.175
Num.Obs.	32	32
RMSE	5.93	5.39
(Intercept)	146.688	631.704
	(12.120)	(88.700)
qsec		-27.174
		(4.946)
R2	0.000	0.502
Num.Obs.	32	32
RMSE	67.48	47.64

```
panels,
shape = "rbind",
gof_map = gm)
```

#### 12.3 String ("cbind"): Automatic spanning column labels

When using the default tinytable backend, we can use a named nested list to automatically add spanning column labels to a table:

```
modelsummary(panels, shape = "cbind")
```

#### 12.4 Renaming estimates and statistics

When statistics are displayed in different columns, users may want control over the statistic names. To do so, we specify the estimate and statistic as named vectors:

	Outcom	e: mpg	Outco	me: hp
	(I)	(II)	(I)	(II)
(Intercept)	20.091	-5.114	146.688	631.704
	(1.065)	(10.030)	(12.120)	(88.700)
qsec		1.412		-27.174
		(0.559)		(4.946)
Num.Obs.	32	32	32	32
R2	0.000	0.175	0.000	0.502
R2 Adj.	0.000	0.148	0.000	0.485
AIC	208.8	204.6	364.4	344.1
BIC	211.7	209.0	367.3	348.5
Log.Lik.	-102.378	-99.294	-180.186	-169.045
F		6.377		30.190
RMSE	5.93	5.39	67.48	47.64

```
mod <- lm(mpg ~ factor(cyl), mtcars)

modelsummary(
    mod,
    estimate = c("$\\hat{\\beta}$" = "estimate"),
    statistic = c("$\\sigma_{\\hat{\\beta}}$" = "std.error"),
    shape = term ~ model + statistic)</pre>
```

#### 12.4.1 fixest

The fixest package offers powerful tools to estimate multiple models using a concise syntax. fixest functions are also convenient because they return named lists of models which are easy to subset and manipulate using standard R functions like grepl.

For example, to introduce regressors in stepwise fashion, and to estimate models on different subsets of the data, we can do:

```
# estimate 4 models
library(fixest)
mod <- feols(
  c(hp, mpg) ~ csw(qsec, drat) | gear,</pre>
```

		(1)
	$\overline{{\hat s}\hat {\beta}}$	$\simeq \$
(Intercept)	26.664	0.972
factor(cyl)6	-6.921	1.558
factor(cyl)8	-11.564	1.299
Num.Obs.	32	
R2	0.732	
R2 Adj.	0.714	
AIC	170.6	
BIC	176.4	
Log.Lik.	-81.282	
F	39.698	
RMSE	3.07	

```
data = mtcars)

# select models with different outcome variables
panels <- list(
    "Miles per gallon" = mod[grepl("mpg", names(mod))],
    "Horsepower" = mod[grepl("hp", names(mod))]
)

modelsummary(
   panels,
   shape = "rcollapse",
   gof_omit = "IC|R2")</pre>
```

We can use all the typical extension systems to add information, such as the mean of the dependent variable:

```
glance_custom.fixest <- function(x, ...) {
  dv <- insight::get_response(x)
  dv <- sprintf("%.2f", mean(dv, na.rm = TRUE))
  data.table::data.table(`Mean of DV` = dv)
}

modelsummary(</pre>
```

	(1)	(2)
qsec	1.436	1.519
	(0.594)	(0.529)
$\operatorname{drat}$		5.765
		(2.381)
RMSE	4.03	3.67
qsec	-22.175	-22.676
	(12.762)	(13.004)
drat		-35.106
		(28.509)
RMSE	40.45	39.14
Num.Obs.	32	32
Std.Errors	by: gear	by: gear
FE: gear	X	X

```
panels,
shape = "rcollapse",
gof_omit = "IC|R2")
```

```
rm("glance_custom.fixest")
```

## 13 align: column alignment

By default, modelsummary will align the first column (with coefficient names) to the left, and will center the results columns. To change this default, you can use the align argument, which accepts a string of the same length as the number of columns:

```
modelsummary(models, align="lrrrrr")
```

Users who produce PDF documents using Rmarkdown or LaTeX can also align values on the decimal dot by using the character "d" in the align argument:

```
modelsummary(models, align="lddddd")
```

	(1)	(2)
qsec	1.436	1.519
	(0.594)	(0.529)
$\operatorname{drat}$		5.765
		(2.381)
RMSE	4.03	3.67
Mean of DV	20.09	20.09
qsec	-22.175	-22.676
	(12.762)	(13.004)
$\operatorname{drat}$		-35.106
		(28.509)
RMSE	40.45	39.14
Mean of DV	146.69	146.69
Num.Obs.	32	32
Std.Errors	by: gear	by: gear
FE: gear	X	X

For the table produced by this code to compile, users must include the following code in their LaTeX preamble:

```
\usepackage{booktabs}
\usepackage{siunitx}
\newcolumntype{d}{S[input-symbols = ()]}
```

### 14 notes: footers

Add notes to the bottom of your table:

## 15 title: captions

You can add a title to your table as follows:

```
modelsummary(models, title = 'This is a title for my table.')
```

#### 16 add\_rows

Use the add\_rows argument to add rows manually to a table. For example, let's say you estimate two models with a factor variables and you want to insert (a) an empty line to identify the category of reference, and (b) customized information at the bottom of the table:

```
models <- list()
models[['OLS']] <- lm(mpg ~ factor(cyl), mtcars)
models[['Logit']] <- glm(am ~ factor(cyl), mtcars, family = binomial)</pre>
```

We create a data.frame with the same number of columns as the summary table. Then, we define a "position" attribute to specify where the new rows should be inserted in the table. Finally, we pass this data.frame to the add\_rows argument:

#### 17 Math

We can display math in tables by enclosing math expressions in dollars signs and setting the argument escape=FALSE:

```
res <- lm(wt ~ mpg + drat, mtcars)

modelsummary(
  res,
  escape = FALSE,
  title = "Another math expression: $\\sum_{i=1}^n x^-i$",
  coef_rename = c("mpg" = "$\\beta_1$"))</pre>
```

	OLS	Logit
(Intercept)	26.664	0.981
	(0.972)	(0.677)
factor(cyl)4	-	-
factor(cyl)6	-6.921	-1.269
	(1.558)	(1.021)
factor(cyl)8	-11.564	-2.773
	(1.299)	(1.021)
Num.Obs.	32	32
Info	???	XYZ
R2	0.732	
R2 Adj.	0.714	
AIC	170.6	39.9
BIC	176.4	44.3
Log.Lik.	-81.282	-16.967
F	39.698	3.691
RMSE	3.07	0.42

Table 1: Another math expression:  $\sum_{i=1}^n x^-i$ 

(1)
7.035
(0.597)
-0.116
(0.019)
-0.415
(0.218)
32
0.780
0.765
47.9
53.8
-19.951
51.510
0.45

## 18 exponentiate

We can exponentiate their estimates using the exponentiate argument:

```
mod_logit <- glm(am ~ mpg, data = mtcars, family = binomial)
modelsummary(mod_logit, exponentiate = TRUE)</pre>
```

We can also present exponentiated and standard models side by side by using a logical vector:

```
mod_logit <- list(mod_logit, mod_logit)
modelsummary(mod_logit, exponentiate = c(TRUE, FALSE))</pre>
```

## 19 ... ellipsis: Additional arguments

All arguments passed by the user to a modelsummary function are pushed forward in two other functions:

	(1)
(Intercept)	0.001
	(0.003)
mpg	1.359
	(0.156)
Num.Obs.	32
AIC	33.7
BIC	36.6
Log.Lik.	-14.838
F	7.148
RMSE	0.39

	(1)	(2)
(Intercept)	0.001	-6.604
	(0.003)	(2.351)
mpg	1.359	0.307
	(0.156)	(0.115)
Num.Obs.	32	32
AIC	33.7	33.7
BIC	36.6	36.6
Log.Lik.	-14.838	-14.838
F	7.148	7.148
RMSE	0.39	0.39

- 1. The function which extracts model estimates.
  - By default, additional arguments are pushed forward to parameters::parameters and performance::performance. Users can also can also use a different "backend" to extract information from model objects: the broom package. By setting the modelsummary\_get global option, we tell modelsummary to use the easystats/parameters packages instead of broom. With these packages, other arguments are available, such as the metrics argument. Please refer to these package's documentation to details.
- 2. The table-making functions.
  - By default, additional arguments are pushed forward to kableExtra::kbl, but users can use a different table-making function by setting the output argument to a different value such as "gt", "flextable", or "huxtable".
  - See the Appearance vignette for examples.

All arguments passed supported by these functions are thus automatically available directly in modelsummary, modelplot, and the datasummary family of functions.

## 20 Custom appearance

To customize the appearance of tables, modelsummary supports six popular and extremely powerful table-making packages:

- 1. tinytable: https://vincentarelbundock.github.io/tinytable/
- 2. gt: https://gt.rstudio.com
- 3. kableExtra: http://haozhu233.github.io/kableExtra
- 4. huxtable: https://hughjonesd.github.io/huxtable/
- 5. flextable: https://davidgohel.github.io/flextable/
- 6. DT: https://rstudio.github.io/DT

The "customizing the look of your tables" vignette shows examples for all 4 packages.

# 21 Supported models

modelsummary automatically supports all the models supported by the tidy function of the broom package or the parameters function of the parameters package. The list of supported models is rapidly expanding. At the moment, it covers the following model classes:

```
supported models()
```

[1] "aareg" "acf" [3] "afex\_aov" "AKP"

[5] "anova" "Anova.mlm"

[7] "anova.rms" "aov" [9] "aovlist" "Arima" [11] "averaging" "bamlss" [13] "bayesQR" "bcplm" [15] "befa" "betamfx" [17] "betaor" "betareg" [19] "BFBayesFactor" "bfsl" [21] "BGGM" "bifeAPEs" "binDesign"

[23] "biglm" "binDesig [25] "binWidth" "blavaan" [27] "blrm" "boot" [29] "bootstrap\_model" "bracl"

[31] "brmsfit" "brmultinom"

[33] "btergm" "cch"
[35] "censReg" "cgam"
[37] "character" "cld"
[39] "clm" "clm2"
[41] "clmm" "clmm2"

[43] "coeftest" "comparisons"
[45] "confint.glht" "confusionMatrix"

[47] "coxph" "cpglmm"
[49] "crr" "cv.glmnet"
[51] "data.frame" "dbscan"

[57] "DirichletRegModel" "dist"
[59] "draws" "drc"
[61] "durbinWatsonTest" "emm\_list"

[63] "emmeans" "emmeans\_summary"

[65] "emmGrid" "epi.2by2"

[67] "ergm" "fa"
[69] "fa.ci" "factanal"
[71] "FAMD" "feglm"
[73] "felm" "fitdistr"

[75] "fixest" "fixest\_multi"

 [77] "flac"
 "flic"

 [79] "ftable"
 "gam"

 [81] "Gam"
 "gamlss"

 [83] "gamm"
 "garch"

 [85] "geeglm"
 "ggeffects"

```
[87] "glht"
                                   "glimML"
 [89] "glm"
                                   "glmm"
 [91] "glmmTMB"
                                   "glmnet"
[93] "glmrob"
                                   "glmRob"
 [95] "glmx"
                                   "gmm"
 [97] "hclust"
                                   "hdbscan"
 [99] "hglm"
                                   "hkmeans"
                                   "htest"
[101] "HLfit"
[103] "hurdle"
                                   "hypotheses"
[105] "irlba"
                                   "ivFixed"
[107] "ivprobit"
                                   "ivreg"
[109] "kappa"
                                   "kde"
[111] "Kendall"
                                   "kmeans"
[113] "lavaan"
                                   "leveneTest"
[115] "Line"
                                   "Lines"
                                   "lm"
[117] "list"
[119] "lm_robust"
                                   "lm.beta"
[121] "lme"
                                   "lmodel2"
[123] "lmrob"
                                   "lmRob"
[125] "logical"
                                   "logistf"
[127] "logitmfx"
                                   "logitor"
[129] "lqm"
                                   "lqmm"
[131] "lsmobj"
                                   "manova"
[133] "maov"
                                   "map"
[135] "marginaleffects"
                                   "marginalmeans"
[137] "margins"
                                   "maxim"
[139] "maxLik"
                                   "mblogit"
[141] "Mclust"
                                   "mcmc"
[143] "mcmc.list"
                                   "MCMCglmm"
[145] "mcp1"
                                   "mcp2"
                                   "mediate"
[147] "med1way"
[149] "merMod"
                                   "merModList"
[151] "meta_bma"
                                   "meta_fixed"
[153] "meta_random"
                                   "metaplus"
[155] "mfx"
                                   "mhurdle"
                                   "mira"
[157] "mipo"
[159] "mixed"
                                   "MixMod"
[161] "mixor"
                                   "mjoint"
                                   "mle2"
[163] "mle"
[165] "mlm"
                                   "mlogit"
[167] "mmrm"
                                   "mmrm_fit"
[169] "mmrm_tmb"
                                   "model_fit"
[171] "model_parameters"
                                   "muhaz"
```

[173] "multinom" "mvord" [175] "negbin" "negbinirr" [177] "negbinmfx" "nestedLogit" [179] "nlrq" "nls" [181] "NULL" "numeric" [183] "omega" "onesampb" [185] "optim" "orcutt" [187] "osrt" "pairwise.htest" [189] "pam" "parameters\_efa" "PCA" [191] "parameters\_pca" [193] "pgmm" "plm" [195] "PMCMR" "poissonirr" [197] "poissonmfx" "poLCA" [199] "polr" "Polygon" [201] "Polygons" "power.htest" [203] "prcomp" "predictions" [205] "principal" "probitmfx" [207] "pvclust" "pyears" [209] "rcorr" "ref.grid" [211] "regsubsets" "ridgelm" [213] "rlm" "rlmerMod" [215] "rma" "robtab" "rq" [217] "roc" [219] "rqs" "rqss" [221] "sarlm" "Sarlm" [223] "scam" "selection" [225] "sem" "SemiParBIV" [227] "slopes" "SpatialLinesDataFrame" [229] "SpatialPolygons" "SpatialPolygonsDataFrame" [231] "spec" "speedglm" [233] "speedlm" "stanfit" "stanreg" [235] "stanmvreg" [237] "summary\_emm" "summary.glht" [239] "summary.lm" "summary.plm" [241] "summaryDefault" "survdiff" "survfit" [243] "survexp" "svd" [245] "survreg" [247] "svyglm" "svyolr" [249] "svytable" "systemfit" [251] "t1way" "table" [253] "tobit" "trendPMCMR" "ts" [255] "trimcibt" [257] "TukeyHSD" "varest"

```
[259] "vgam" "wbgee"
[261] "wbm" "wmcpAKP"
[263] "xyz" "yuen"
[265] "zcpglm" "zerocount"
[267] "zeroinfl" "zoo"
```

To see if a given model is supported, you can fit it, and then call this function:

```
get_estimates(model)
```

If this function does not return a valid output, you can easily (really!!) add your own support. See the next section for a tutorial. If you do this, you may consider opening an issue on the Github website of the broom package: https://github.com/tidymodels/broom/issues

## 22 Rmarkdown, Quarto, Org-Mode

#### 22.1 Rmarkdown

You can use modelsummary to insert tables into dynamic documents with knitr or Rmarkdown. This minimal .Rmd file can produce tables in PDF, HTML, or RTF documents:

• minimal.Rmd

This .Rmd file shows illustrates how to use table numbering and cross-references to produce PDF documents using bookdown:

• cross references.Rmd

This .Rmd file shows how to customize tables in PDF and HTML files using gt and kableExtra functions:

• appearance.Rmd

#### 22.2 Quarto

Quarto is an open source publishing system built on top of Pandoc. It was designed as a "successor" to Rmarkdown, and includes useful features for technical writing, such as built-in support for cross-references. modelsummary works automatically with Quarto. This is a minimal document with cross-references which should render automatically to PDF, HTML, and more:

```
format: pdf
title: Example
---

Otbl-mtcars shows that cars with high horse power get low miles per gallon.

```{r}

#| label: tbl-mtcars

#| tbl-cap: "Horse Powers vs. Miles per Gallon"
library(modelsummary)

mod <- lm(mpg ~ hp, mtcars)

modelsummary(mod)

```
```

## 22.3 Emacs Org-Mode

#+END\_SRC

You can use modelsummary to insert tables into Emacs Org-Mode documents, which can be exported to a variety of formats, including HTML and PDF (via LaTeX). As with anything Emacs-related, there are many ways to achieve the outcomes you want. Here is one example of an Org-Mode document which can automatically export tables to HTML and PDF without manual tweaks:

```
#+PROPERTY: header-args:R :var orgbackend=(prin1-to-string org-export-current-backend)
#+MACRO: Rtable (eval (concat "#+header: :results output " (prin1-to-string org-export-current)
{{Rtable}}}
#+BEGIN_SRC R :exports both
library(modelsummary)
options(modelsummary_factory_default = orgbackend)

mod = lm(hp ~ mpg, data = mtcars)

modelsummary(mod)
```

The first line tells Org-mode to assign a variable called **orgbackend**. This variable will be accessible by the R session, and will be equal to "html" or "latex", depending on the export format.

The second line creates an Org macro which we will use to automatically add useful information to the header of source blocks. For instance, when we export to HTML, the macro will expand

to :results output html. This tells Org-Mode to insert the last printed output from the R session, and to treat it as raw HTML.

The {{Rtable}} call expands the macro to add information to the header of the block that follows.

#+BEGIN\_SRC R :exports both says that we want to print both the original code and the output (:exports results would omit the code, for example).

Finally, options (modelsummary\_factory\_default=orgbackend uses the variable we defined to set the default output format. That way, we don't have to use the output argument every time.

One potentially issue to keep in mind is that the code above extracts the printout from the R console. However, when we customize tables with kableExtra or gt functions, those functions do not always return printed raw HTML or LaTeX code. Sometimes, it can be necessary to add a call to cat at the end of a table customization pipeline. For example:

```
{{Rtable}}}
#+BEGIN_SRC R :exports both
library(modelsummary)
library(kableExtra)

mod = lm(hp ~ mpg, data = mtcars)

modelsummary(mod, output = orgbackend) %>%
   row_spec(1, background = "pink") %>%
   cat()
#+END_SRC
```

## 23 Global options

Users can change the default behavior of modelsummary by setting global options.

Omit the note at the bottom of the table with significance threshold:

```
options("modelsummary_stars_note" = FALSE)
```

Change the default output format:

```
options(modelsummary_factory_default = "latex")
options(modelsummary_factory_default = "gt")
```

Change the backend packages that modelsummary uses to create tables in different output formats:

```
options(modelsummary_factory_html = 'kableExtra')
options(modelsummary_factory_latex = 'flextable')
options(modelsummary_factory_word = 'huxtable')
options(modelsummary_factory_png = 'gt')
```

Change the packages that modelsummary uses to extract information from models:

```
# tidymodels: broom
options(modelsummary_get = "broom")

# easystats: performance + parameters
options(modelsummary_get = "easystats")
```

The appearance vignette shows how to set "themes" for your tables using the modelsummary\_theme\_gt, modelsummary\_theme\_kableExtra, modelsummary\_theme\_flextable and modelsummary\_theme\_huxtable global options. For example:

```
library(gt)

# The ... ellipsis is required!
custom_theme <- function(x, ...) {
    x %>% gt::opt_row_striping(row_striping = TRUE)
}
options("modelsummary_theme_gt" = custom_theme)

mod <- lm(mpg ~ hp + drat, mtcars)
modelsummary(mod, output = "gt")</pre>
```

#### 24 Case studies

### 24.1 Reference categories

When estimating models with factor regressors, R will typically omit the reference category. We can include an empty row for the reference level automatically by using the include\_reference argument. This argument is supplied by the parameters package, which modelsummary uses behind the scenes to extract model estimates. We can supply it directly to modelsummary(), which will then pass the argument forward automatically:

| (1)     |
|---------|
| 26.664  |
| (0.972) |
| -       |
| -6.921  |
| (1.558) |
| -11.564 |
| (1.299) |
| 32      |
| 0.732   |
| 0.714   |
| 170.6   |
| 176.4   |
| -81.282 |
| 39.698  |
| 3.07    |
|         |

```
tmp <- transform(mtcars, cyl = factor(cyl))
mod <- lm(mpg ~ cyl, tmp)
modelsummary(mod, include_reference = TRUE)</pre>
```

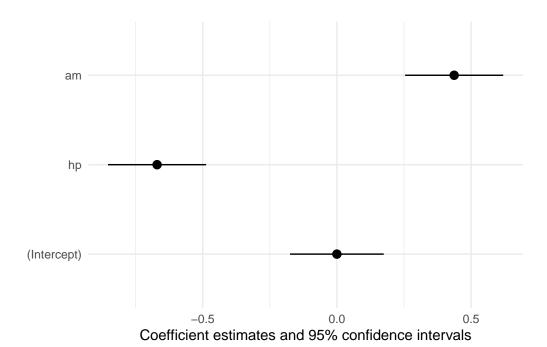
#### 24.2 Standardization

In some cases, it is useful to standardize coefficients before reporting them. modelsummary extracts coefficients from model objects using the parameters package, and that package offers several options for standardization: https://easystats.github.io/parameters/reference/model\_parameters.defaul

We can pass the standardize argument directly to modelsummary or modelplot, and that argument will be forwarded to parameters. For example to refit the model on standardized data and plot the results, we can do:

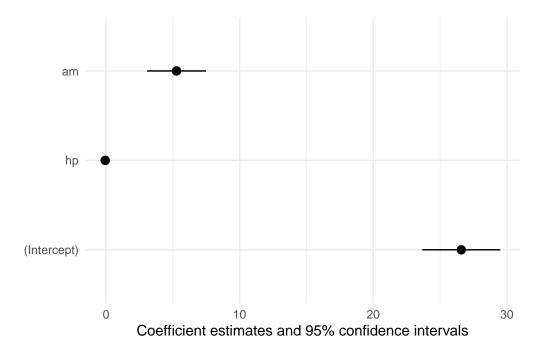
```
library(modelsummary)
mod <- lm(mpg ~ hp + am, data = mtcars)
modelplot(mod, standardize = "refit")</pre>
```

Warning in  $ggplot2::geom_pointrange(ggplot2::aes(y = term, x = estimate, : Ignoring unknown parameters: `standardize`$ 



Compare to the unstandardized plot:

modelplot(mod)



### 24.3 Subgroup estimation with nest\_by

Sometimes, it is useful to estimate multiple regression models on subsets of the data. To do this efficiently, we can use the nest\_by function from the dplyr package. Then, estimate the models with lm, extract them and name them with pull, and finally summarize them with modelsummary:

```
library(tidyverse)

mtcars %>%
    nest_by(cyl) %>%
    mutate(models = list(lm(mpg ~ hp, data))) %>%
    pull(models, name = cyl) %>%
    modelsummary
```

## 24.4 Statistics in separate columns instead of one over the other

In somes cases, you may want to display statistics in separate columns instead of one over the other. It is easy to achieve this outcome by using the estimate argument. This argument accepts a vector of values, one for each of the models we are trying to summarize. If we want

|             | 4       | 6       | 8       |
|-------------|---------|---------|---------|
| (Intercept) | 35.983  | 20.674  | 18.080  |
|             | (5.201) | (3.304) | (2.988) |
| hp          | -0.113  | -0.008  | -0.014  |
|             | (0.061) | (0.027) | (0.014) |
| Num.Obs.    | 11      | 7       | 14      |
| R2          | 0.274   | 0.016   | 0.080   |
| R2 Adj.     | 0.193   | -0.181  | 0.004   |
| AIC         | 65.8    | 29.9    | 69.8    |
| BIC         | 67.0    | 29.7    | 71.8    |
| Log.Lik.    | -29.891 | -11.954 | -31.920 |
| RMSE        | 3.66    | 1.33    | 2.37    |

to include estimates and standard errors in separate columns, all we need to do is repeat a model, but request different statistics. For example,

```
library(modelsummary)
library(kableExtra)
```

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

group\_rows

|             | Мс     | odel A    | Mo     | odel B    |
|-------------|--------|-----------|--------|-----------|
|             | Coef.  | Std.Error | Coef.  | Std.Error |
| (Intercept) | 30.099 | 1.634     | 10.790 | 5.078     |
| hp          | -0.068 | 0.010     | -0.052 | 0.009     |
| drat        |        |           | 4.698  | 1.192     |

|             | Coef.  | Std.Error | Coef.  | Std.Error |
|-------------|--------|-----------|--------|-----------|
| (Intercept) | 30.099 | 1.634     | 10.790 | 5.078     |
| hp          | -0.068 | 0.010     | -0.052 | 0.009     |
| drat        |        |           | 4.698  | 1.192     |

This can be automated using a simple function:

### 24.5 Bootstrap

Users often want to use estimates or standard errors that have been obtained using a custom strategy. To achieve this in an automated and replicable way, it can be useful to use the

tidy\_custom strategy described above in the "Cutomizing Existing Models" section.

For example, we can use the modelr package to draw 500 resamples of a dataset, and compute bootstrap standard errors by taking the standard deviation of estimates computed in all of those resampled datasets. To do this, we defined tidy\_custom.lm function that will automatically bootstrap any lm model supplied to modelsummary, and replace the values in the table automatically.

Note that the tidy\_custom\_lm returns a data.frame with 3 columns: term, estimate, and std.error:

```
library("modelsummary")
library("broom")
library("tidyverse")
library("modelr")
tidy_custom.lm <- function(x, ...) {</pre>
  # extract data from the model
  model.frame(x) %>%
    # draw 500 bootstrap resamples
    modelr::bootstrap(n = 500) \%>\%
    # estimate the model 500 times
    mutate(results = map(strap, ~ update(x, data = .))) %>%
    # extract results using `broom::tidy`
    mutate(results = map(results, tidy)) %>%
    # unnest and summarize
    unnest(results) %>%
    group_by(term) %>%
    summarize(std.error = sd(estimate),
              estimate = mean(estimate))
}
mod = list(
  lm(hp ~ mpg, mtcars) ,
  lm(hp ~ mpg + drat, mtcars))
modelsummary(mod)
```

#### 24.6 fixest: Fixed effects and instrumental variable regression

One common use-case for glance\_custom is to include additional goodness-of-fit statistics. For example, in an instrumental variable estimation computed by the fixest package, we

|                       | (1)      | (2)      |
|-----------------------|----------|----------|
| (Intercept)           | 330.716  | 280.728  |
|                       | (31.489) | (42.778) |
| mpg                   | -9.153   | -10.085  |
|                       | (1.432)  | (2.215)  |
| $\operatorname{drat}$ |          | 18.774   |
|                       |          | (20.658) |
| Num.Obs.              | 32       | 32       |
| R2                    | 0.602    | 0.614    |
| R2 Adj.               | 0.589    | 0.588    |
| AIC                   | 336.9    | 337.9    |
| BIC                   | 341.3    | 343.7    |
| Log.Lik.              | -165.428 | -164.940 |
| F                     | 45.460   | 23.100   |
| RMSE                  | 42.55    | 41.91    |

may want to include an IV-Wald statistic for the first-stage regression of each endogenous regressor:

```
library(fixest)
library(tidyverse)

# create a toy dataset
base <- iris
names(base) <- c("y", "x1", "x_endo_1", "x_inst_1", "fe")
base$x_inst_2 <- 0.2 * base$y + 0.2 * base$x_endo_1 + rnorm(150, sd = 0.5)
base$x_endo_2 <- 0.2 * base$y - 0.2 * base$x_inst_1 + rnorm(150, sd = 0.5)

# estimate an instrumental variable model
mod <- feols(y ~ x1 | fe | x_endo_1 + x_endo_2 ~ x_inst_1 + x_inst_2, base)

# custom extractor function returns a one-row data.frame (or tibble)
glance_custom.fixest <- function(x) {
  tibble(
    "Wald (x_endo_1)" = fitstat(x, "ivwald")[[1]]$stat,
    "Wald (x_endo_2)" = fitstat(x, "ivwald")[[2]]$stat
)</pre>
```

|                       | (1)              |
|-----------------------|------------------|
| fit_x_endo_1          | 0.475            |
|                       | (0.160)          |
| $fit\_x\_endo\_2$     | 0.810            |
|                       | (1.409)          |
| x1                    | 0.531            |
|                       | (0.141)          |
| Num.Obs.              | 150              |
| R2                    | 0.649            |
| R2 Adj.               | 0.636            |
| R2 Within             | 0.079            |
| R2 Within Adj.        | 0.059            |
| AIC                   | 223.2            |
| BIC                   | 241.2            |
| RMSE                  | 0.49             |
| Std.Errors            | by: fe           |
| FE: fe                | X                |
| $Wald \ (x\_endo\_1)$ | 24.0483473138048 |
| Wald (x_endo_2)       | 1.01931209406748 |

```
# draw table
modelsummary(mod)

rm("glance_custom.fixest")
```

## 24.7 Multiple imputation

 ${\tt modelsummary}$  can pool and display analyses on several datasets imputed using the  ${\tt mice}$  or  ${\tt Amelia}$  packages. This code illustrates how:

```
library(mice)
library(Amelia)
```

```
library(modelsummary)
# Download data from `Rdatasets`
url <- 'https://vincentarelbundock.github.io/Rdatasets/csv/HistData/Guerry.csv'
dat <- read.csv(url)[, c('Clergy', 'Commerce', 'Literacy')]</pre>
# Insert missing values
dat$Clergy[sample(1:nrow(dat), 10)] <- NA</pre>
dat$Commerce[sample(1:nrow(dat), 10)] <- NA</pre>
dat$Literacy[sample(1:nrow(dat), 10)] <- NA</pre>
# Impute with `mice` and `Amelia`
dat_mice <- mice(dat, m = 5, printFlag = FALSE)</pre>
dat_amelia \leftarrow amelia(dat, m = 5, p2s = 0)$imputations
# Estimate models
mod <- list()</pre>
mod[['Listwise deletion']] <- lm(Clergy ~ Literacy + Commerce, dat)</pre>
mod[['Mice']] <- with(dat_mice, lm(Clergy ~ Literacy + Commerce))</pre>
mod[['Amelia']] <- lapply(dat_amelia, function(x) lm(Clergy ~ Literacy + Commerce, x))</pre>
# Pool results
mod[['Mice']] <- mice::pool(mod[['Mice']])</pre>
mod[['Amelia']] <- mice::pool(mod[['Amelia']])</pre>
# Summarize
modelsummary(mod)
```

## 25 Table-making packages

The table-making backends supported by modelsummary have overlapping capabilities (e.g., several of them can produce HTML tables). These are the default packages used for different outputs:

kableExtra:

- HTML
- LaTeX / PDF

flextable:

|             | Listwise deletion | Mice     | Amelia   |
|-------------|-------------------|----------|----------|
| (Intercept) | 89.092            | 93.664   | 84.457   |
|             | (13.796)          | (13.254) | (14.408) |
| Literacy    | -0.561            | -0.701   | -0.576   |
|             | (0.223)           | (0.208)  | (0.227)  |
| Commerce    | -0.526            | -0.473   | -0.422   |
|             | (0.147)           | (0.140)  | (0.155)  |
| Num.Obs.    | 58                | 86       | 86       |
| Num.Imp.    |                   | 5        | 5        |
| R2          | 0.192             | 0.177    | 0.145    |
| R2 Adj.     | 0.163             | 0.158    | 0.124    |
| AIC         | 535.2             |          |          |
| BIC         | 543.5             |          |          |
| Log.Lik.    | -263.606          |          |          |
| RMSE        | 22.78             |          |          |

- $\bullet$  Word
- Powerpoint

## gt:

- jpg
- png

You can modify these defaults by setting global options such as:

```
options(modelsummary_factory_html = "kableExtra")
options(modelsummary_factory_latex = "gt")
options(modelsummary_factory_word = "huxtable")
options(modelsummary_factory_png = "gt")
```

## 26 FAQ

## 26.1 Stack Overflow is useful

- Standardized coefficients
- Row group labels

- Customizing Word tables
- How to add p values to datasummary\_correlation

### 26.2 Where can I get help?

First, please read the documentation in ?modelsummary and on the modelsummary website. The website includes dozens of worked examples and a lot of detailed explanation.

Second, try to use the [modelsummary] tag on StackOverflow.

Third, if you think you found a bug or have a feature request, please file it on the Github issue tracker:

## 26.3 How can I add or modify statistics in a table?

See the detailed documentation in the "Adding and Customizing Models" section of the modelsummary website.

#### 26.4 How does model summary extract estimates and goodness-of-fit statistics?

A modelsummary table is divided in two parts: "Estimates" (top of the table) and "Goodness-of-fit" (bottom of the table). To populate those two parts, modelsummary tries using the broom, parameters and performance packages in sequence.

#### Estimates:

- 1. Try the broom::tidy function to see if that package supports this model type, or if the user defined a custom tidy function in their global environment. If this fails...
- 2. Try the parameters::model\_parameters function to see if the parameters package supports this model type.

#### Goodness-of-fit:

- 1. Try the performance::model\_performance function to see if the performance package supports this model type.
- 2. Try the broom::glance function to see if that package supports this model type, or if the user defined a custom glance function in their global environment. If this fails...

You can change the order in which those steps are executed by setting a global option:

```
# tidymodels: broom
options(modelsummary_get = "broom")

# easystats: performance + parameters
options(modelsummary_get = "easystats")
```

If all of this fails, modelsummary will return an error message.

If you have problems with a model object, you can often diagnose the problem by running the following commands from a clean R session:

```
# see if parameters and performance support your model type
library(parameters)
library(performance)
model_parameters(model)
model_performance(model)

# see if broom supports your model type
library(broom)
tidy(model)
glance(model)

# see if broom.mixed supports your model type
library(broom.mixed)
tidy(model)
glance(model)
```

If none of these options work, you can create your own tidy and glance methods, as described in the Adding new models section.

If one of the extractor functions does not work well or takes too long to process, you can define a new "custom" model class and choose your own extractors, as described in the Adding new models section.

#### 26.5 How can I speed up modelsummary?

The modelsummary function, by itself, is *not* slow: it should only take a couple seconds to produce a table in any output format. However, sometimes it can be computationally expensive (and long) to extract estimates and to compute goodness-of-fit statistics for your model.

The main options to speed up modelsummary are:

1. Set gof\_map=NA to avoid computing expensive goodness-of-fit statistics.

- 2. Use the easystats extractor functions and the metrics argument to avoid computing expensive statistics (see below for an example).
- 3. Use parallel computation if you are summarizing multiple models. See the "Parallel computation" section in the ?modelsummary documentation.

To diagnose the slowdown and find the bottleneck, you can try to benchmark the various extractor functions:

```
library(tictoc)

data(trade)
mod <- lm(mpg ~ hp + drat, mtcars)

tic("tidy")
x <- broom::tidy(mod)
toc()</pre>
```

tidy: 0.003 sec elapsed

```
tic("glance")
x <- broom::glance(mod)
toc()</pre>
```

glance: 0.004 sec elapsed

```
tic("parameters")
x <- parameters::parameters(mod)
toc()</pre>
```

parameters: 0.022 sec elapsed

```
tic("performance")
x <- performance::performance(mod)
toc()</pre>
```

performance: 0.012 sec elapsed

In my experience, the main bottleneck tends to be computing goodness-of-fit statistics. The performance extractor allows users to specify a metrics argument to select a subset of GOF to include. Using this can speedup things considerably.

We call modelsummary with the metrics argument:

|             | (1)     |
|-------------|---------|
| (Intercept) | 10.790  |
|             | (5.078) |
| hp          | -0.052  |
|             | (0.009) |
| drat        | 4.698   |
|             | (1.192) |
| Num.Obs.    | 32      |
| R2          | 0.741   |
| R2 Adj.     | 0.723   |
| AIC         | 169.5   |
| BIC         | 175.4   |
| Log.Lik.    | -80.752 |
| F           | 41.522  |

modelsummary(mod, metrics = "rmse")

## 26.6 Escaped LaTeX characters

Sometimes, users want to include raw LaTeX commands in their tables, such as coefficient names including math mode: Apple \$\times\$ Orange. The result of these attempts is often a weird string such as: \\$\textbackslash{}\times\\$ instead of proper LaTeX-rendered characters.

The source of the problem is that kableExtra, default table-making package in modelsummary, automatically escapes weird characters to make sure that your tables compile properly in LaTeX. To avoid this, we need to pass the escape=FALSE to modelsummary:

```
modelsummary(mod, escape = FALSE)
```

## 26.7 Bayesian models

Many bayesian models are supported out-of-the-box, including those produced by the rstanarm and brms packages. The statistics available for bayesian models are slightly different than those available for most frequentist models. Users can call get\_estimates to see what is available:

#### library(rstanarm)

This is rstanarm version 2.32.1

- See https://mc-stan.org/rstanarm/articles/priors for changes to default priors!
- Default priors may change, so it's safest to specify priors, even if equivalent to the def
- For execution on a local, multicore CPU with excess RAM we recommend calling

```
options(mc.cores = parallel::detectCores())
```

Attaching package: 'rstanarm'

The following object is masked from 'package:fixest':

se

```
mod <- stan_glm(am ~ hp + drat, data = mtcars)</pre>
```

### get\_estimates(mod)

```
mad conf.level
                   estimate
                                                       conf.low
                                                                  conf.high
1 (Intercept) -2.2194596943 0.608059109
                                              0.95 -3.434491263 -1.06593762
2
          hp 0.0007002131 0.001088948
                                              0.95 -0.001423932 0.00281773
         drat 0.7025778679 0.144207613
                                              0.95 0.435730950 0.99455983
 prior.distribution prior.location prior.scale group std.error statistic
                            0.40625 1.24747729
                                                             NA
                                                                       NA
1
              normal
2
              normal
                            0.00000 0.01819465
                                                             NA
                                                                       NA
3
              normal
                            0.00000 2.33313429
                                                             NA
                                                                       NA
 p.value
       NA
2
       NA
3
       NA
```

This shows that there is no std.error column, but that there is a mad statistic (mean absolute deviation). So we can do:

|             | (1)     |
|-------------|---------|
| (Intercept) | -2.219  |
|             | (0.608) |
| hp          | 0.001   |
|             | (0.001) |
| drat        | 0.703   |
|             | (0.144) |
| Num.Obs.    | 32      |
| R2          | 0.502   |
| R2 Adj.     | 0.433   |
| Log.Lik.    | -12.047 |
| ELPD        | -15.1   |
| ELPD s.e.   | 3.1     |
| LOOIC       | 30.2    |
| LOOIC s.e.  | 6.2     |
| WAIC        | 30.0    |
| RMSE        | 0.34    |

#### modelsummary(mod, statistic = "mad")

#### Warning:

`modelsummary` uses the `performance` package to extract goodness-of-fit statistics from models of this class. You can specify the statistics you wish to compute by supplying a `metrics` argument to `modelsummary`, which will then push it forward to `performance`. Acceptable values are: "all", "common", "none", or a character vector of metrics names. For example: `modelsummary(mod, metrics = c("RMSE", "R2")` Note that some metrics are computationally expensive. See `?performance::performance` for details.

This warning appears once per session.

As noted in the modelsummary() documentation, model results are extracted using the parameters package. Users can pass additional arguments to modelsummary(), which will then push forward those arguments to the parameters::parameters function to change the results. For example, the parameters documentation for bayesian models shows that there is a centrality argument, which allows users to report the mean and standard deviation of the posterior distribution, instead of the median and MAD:

| (1)     |
|---------|
| -2.242  |
| (0.603) |
| 0.001   |
| (0.001) |
| 0.707   |
| (0.143) |
| 32      |
| 0.502   |
| 0.433   |
| -12.047 |
| -15.1   |
| 3.1     |
| 30.2    |
| 6.2     |
| 30.0    |
| 0.34    |
|         |

## get\_estimates(mod, centrality = "mean")

```
term
                  estimate
                               std.dev conf.level
                                                      conf.low
                                                                 conf.high
1 (Intercept) -2.2415320842 0.603076209
                                        0.95 -3.434491263 -1.06593762
          hp 0.0007209036 0.001083671
                                             0.95 -0.001423932 0.00281773
2
        drat 0.7066846030 0.142648838
                                             0.95 0.435730950 0.99455983
 prior.distribution prior.location prior.scale group std.error statistic
              normal
                           0.40625 1.24747729
                                                                      NA
                                                            NA
                           0.00000 0.01819465
                                                            NA
2
             normal
                                                                      NA
3
             normal
                           0.00000 2.33313429
                                                            NA
                                                                      NA
 p.value
1
      NA
2
      NA
3
      NA
```

```
modelsummary(mod, statistic = "std.dev", centrality = "mean")
```

We can also get additional test statistics using the test argument:

### get\_estimates(mod, test = c("pd", "rope"))

```
mad conf.level
                                                                conf.high
         term
                   estimate
                                                       conf.low
1 (Intercept) -2.2194596943 0.608059109
                                             0.95 -3.434491263 -1.06593762
          hp 0.0007002131 0.001088948
                                             0.95 -0.001423932 0.00281773
2
3
         drat 0.7025778679 0.144207613
                                             0.95 0.435730950 0.99455983
      pd rope.percentage prior.distribution prior.location prior.scale group
                                                    0.40625 1.24747729
1 0.99975
                        0
                                     normal
2 0.74825
                        1
                                     normal
                                                    0.00000 0.01819465
3 1.00000
                                                    0.00000 2.33313429
                        0
                                     normal
 std.error statistic p.value
1
        NA
                  NA
2
                  NA
         NA
                          NA
3
        NA
                  NA
                          NA
```