

Package ‘equatiomatic’

January 30, 2021

Title Transform Models into 'LaTeX' Equations

Version 0.2.0

Description The goal of 'equatiomatic' is to reduce the pain associated with writing 'LaTeX' formulas from fitted models. The primary function of the package, `extract_eq()`, takes a fitted model object as its input and returns the corresponding 'LaTeX' code for the model.

License MIT + file LICENSE

Depends R (>= 3.3.0)

Imports broom (>= 0.7.0), stats, knitr, broom.mixed, utils

Suggests texPreview, testthat (>= 3.0.0), MASS, ordinal, lme4, rmarkdown, covr, forecast (>= 8.13), latex2exp (>= 0.4.0), ggplot2 (>= 3.3.2)

URL <https://github.com/datalorax/equatiomatic>

BugReports <https://github.com/datalorax/equatiomatic/issues>

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

VignetteBuilder knitr

Config/testthat/edition 3

NeedsCompilation no

Author Daniel Anderson [aut, cre] (<<https://orcid.org/0000-0003-4699-4680>>),
Andrew Heiss [aut] (<<https://orcid.org/0000-0002-3948-3914>>),
Jay Sumners [aut],
Joshua Rosenberg [ctb] (<<https://orcid.org/0000-0003-2170-0447>>),
Jonathan Sidi [ctb] (<<https://orcid.org/0000-0002-4222-1819>>),
Ellis Hughes [ctb] (<<https://orcid.org/0000-0003-0637-4436>>),
Thomas Fung [ctb] (<<https://orcid.org/0000-0003-2601-0728>>),
Reza Norouzian [ctb] (<<https://orcid.org/0000-0002-0531-6886>>)

Maintainer Daniel Anderson <daniela@uoregon.edu>

Repository CRAN

Date/Publication 2021-01-30 06:20:20 UTC

R topics documented:

extract_eq	2
hsb	4
penguins	5
print.equation	6
simple_ts	6
sim_longitudinal	7
ts_reg_list	8

Index	9
--------------	----------

extract_eq	<i>'LaTeX' code for R models</i>
------------	----------------------------------

Description

Extract the variable names from a model to produce a 'LaTeX' equation, which is output to the screen. Supports any model supported by [broom::tidy](#).

Usage

```
extract_eq(
  model,
  intercept = "alpha",
  greek = "beta",
  raw_tex = FALSE,
  ital_vars = FALSE,
  show_distribution = FALSE,
  wrap = FALSE,
  terms_per_line = 4,
  operator_location = "end",
  align_env = "aligned",
  use_coefs = FALSE,
  coef_digits = 2,
  fix_signs = TRUE,
  mean_separate,
  ...
)
```

Arguments

model	A fitted model
intercept	How should the intercept be displayed? Default is "alpha", but can also accept "beta", in which case the it will be displayed as beta zero.
greek	What notation should be used for coefficients? Currently only accepts "beta" (with plans for future development). Can be used in combination with raw_tex to use any notation, e.g., "\hat{\beta}".

raw_tex	Logical. Is the greek code being passed to denote coefficients raw tex code?
ital_vars	Logical, defaults to FALSE. Should the variable names not be wrapped in the <code>\operatorname{}</code> command?
show_distribution	Logical. When fitting a logistic or probit regression, should the binomial distribution be displayed? Defaults to FALSE.
wrap	Logical, defaults to FALSE. Should the terms on the right-hand side of the equation be split into multiple lines? This is helpful with models with many terms.
terms_per_line	Integer, defaults to 4. The number of right-hand side terms to include per line. Used only when wrap is TRUE.
operator_location	Character, one of “end” (the default) or “start”. When terms are split across multiple lines, they are split at mathematical operators like +. If set to “end”, each line will end with a trailing operator (+ or -). If set to “start”, each line will begin with an operator.
align_env	TeX environment to wrap around equation. Must be one of aligned, aligned*, align, or align*. Defaults to aligned.
use_coefs	Logical, defaults to FALSE. Should the actual model estimates be included in the equation instead of math symbols?
coef_digits	Integer, defaults to 2. The number of decimal places to round to when displaying model estimates.
fix_signs	Logical, defaults to FALSE. If disabled, coefficient estimates that are negative are preceded with a "+" (e.g. $5(x) + -3(z)$). If enabled, the "+ -" is replaced with "-" (e.g. $5(x) -3(z)$).
mean_separate	Currently only support for lmer models. Should the mean structure be inside or separated from the normal distribution? Defaults to NULL, in which case it will become TRUE if there are more than three fixed-effect parameters. If TRUE, the equation will be displayed as, for example, $\text{outcome} \sim N(\mu, \sigma); \mu = \alpha + \beta_1(\text{wave})$. If FALSE, this same equation would be $\text{outcome} \sim N(\alpha + \beta, \sigma)$.
...	Additional arguments (for future development; not currently used).

Value

A character of class “equation”.

Examples

```
# Simple model
mod1 <- lm(mpg ~ cyl + disp, mtcars)
extract_eq(mod1)

# Include all variables
mod2 <- lm(mpg ~ ., mtcars)
extract_eq(mod2)

# Works for categorical variables too, putting levels as subscripts
```

```

mod3 <- lm(body_mass_g ~ bill_length_mm + species, penguins)
extract_eq(mod3)

set.seed(8675309)
d <- data.frame(cat1 = rep(letters[1:3], 100),
                cat2 = rep(LETTERS[1:3], each = 100),
                cont1 = rnorm(300, 100, 1),
                cont2 = rnorm(300, 50, 5),
                out = rnorm(300, 10, 0.5))
mod4 <- lm(out ~ ., d)
extract_eq(mod4)

# Don't italicize terms
extract_eq(mod1, ital_vars = FALSE)

# Wrap equations in an "aligned" environment
extract_eq(mod2, wrap = TRUE)

# Wider equation wrapping
extract_eq(mod2, wrap = TRUE, terms_per_line = 4)

# Include model estimates instead of Greek letters
extract_eq(mod2, wrap = TRUE, terms_per_line = 2, use_coefs = TRUE)

# Don't fix doubled-up "+ -" signs
extract_eq(mod2, wrap = TRUE, terms_per_line = 4, use_coefs = TRUE, fix_signs = FALSE)

# Use other model types, like glm
set.seed(8675309)
d <- data.frame(out = sample(0:1, 100, replace = TRUE),
                cat1 = rep(letters[1:3], 100),
                cat2 = rep(LETTERS[1:3], each = 100),
                cont1 = rnorm(300, 100, 1),
                cont2 = rnorm(300, 50, 5))
mod5 <- glm(out ~ ., data = d, family = binomial(link = "logit"))
extract_eq(mod5, wrap = TRUE)

```

Description

This is the dataset used throughout Raudenbush & Bryk (2002).

Usage

hsb

Format

A tibble with 7185 rows and 8 variables:

sch.id An integer denoting the school identification number. There are 160 unique schools.

math Individual students' math score.

size The number of students in the school.

sector A dummy variable (integer) denoting whether the school is public (sector = 0) or catholic (sector = 1). There are 90 public schools and 70 catholic.

meanses A group-mean centered SES variable at the school level

minority A dummy variable indicating if the student was coded as white (minority = 0) or not (minority = 1).

female A dummy variable indicating if the student was coded as female (female = 1) or not (female = 0).

ses A student-level composite variable indicating the students' socio-economic status.

penguins

Size measurements for adult foraging penguins near Palmer Station, Antarctica

Description

Data originally from [palmerpenguins](#). Includes measurements for penguin species, island in Palmer Archipelago, size (flipper length, body mass, bill dimensions), and sex.

Usage

penguins

Format

A tibble with 344 rows and 8 variables:

species a factor denoting penguin species (Adélie, Chinstrap and Gentoo)

island a factor denoting island in Palmer Archipelago, Antarctica (Biscoe, Dream or Torgersen)

bill_length_mm a number denoting bill length (millimeters)

bill_depth_mm a number denoting bill depth (millimeters)

flipper_length_mm an integer denoting flipper length (millimeters)

body_mass_g an integer denoting body mass (grams)

sex a factor denoting penguin sex (female, male)

year an integer denoting the study year (2007, 2008, or 2009)

Source

Adélie penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Adélie penguins (*Pygoscelis adeliae*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 5. Environmental Data Initiative doi: [10.6073/pasta/98b16d7d563f265cb52372c8ca99e60f](https://doi.org/10.6073/pasta/98b16d7d563f265cb52372c8ca99e60f)

Gentoo penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Gentoo penguin (*Pygoscelis papua*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 5. Environmental Data Initiative doi: [10.6073/pasta/7fca67fb28d56ee2ffa3d9370ebda689](https://doi.org/10.6073/pasta/7fca67fb28d56ee2ffa3d9370ebda689)

Chinstrap penguins: Palmer Station Antarctica LTER and K. Gorman. 2020. Structural size measurements and isotopic signatures of foraging among adult male and female Chinstrap penguin (*Pygoscelis antarcticus*) nesting along the Palmer Archipelago near Palmer Station, 2007-2009 ver 6. Environmental Data Initiative doi: [10.6073/pasta/c14dfcfada8ea13a17536e73eb6fbe9e](https://doi.org/10.6073/pasta/c14dfcfada8ea13a17536e73eb6fbe9e)

Originally published in: Gorman KB, Williams TD, Fraser WR (2014) Ecological Sexual Dimorphism and Environmental Variability within a Community of Antarctic Penguins (Genus *Pygoscelis*). PLoS ONE 9(3): e90081. doi:10.1371/journal.pone.0090081

print.equation *Print 'LaTeX' equations*

Description

Print 'LaTeX' equations built with [extract_eq](#).

Usage

```
## S3 method for class 'equation'
print(x, ...)
```

Arguments

x	'LaTeX' equation built with extract_eq
...	not used

simple_ts *Simple simulated time series data*

Description

Output from `set.seed(42); simple_ts <- ts(rnorm(1000), freq = 4)`. This is included primarily for unit testing.

Usage

```
simple_ts
```

Format

A tibble with 1000 rows and 8 variables:

Qtr1 First quarter simulated values.

Qtr2 Second quarter simulated values.

Qtr3 Third quarter simulated values.

Qtr4 Fourth quarter simulated values.

sim_longitudinal	<i>Simulated longitudinal data</i>
------------------	------------------------------------

Description

Data are simulated to be similar to longitudinal data collected within schools/districts.

Usage

```
sim_longitudinal
```

Format

A tibble with 1000 rows and 8 variables:

sid An integer denoting the individual student. There are 100 students.

school An integer denoting the school. There are 15 schools.

district An integer denoting the school district. There are 5 districts.

group A character variable denoting the instructional level of the student, low, medium, or high.

treatment A factor indicating whether the student received the intervention treatment (0 = no treatment received; 1 = treatment received).

prop_low The proportion of student in the school in the low instructional group.

wave The assessment wave. Each student has nine waves of data collection.

score The individual students' score at the given wave.

`ts_reg_list`*Simulated data for time-series regression*

Description

Output from `set.seed(42); ts_reg_list <- list(x1 = rnorm(1000), x2 = rnorm(1000), ts_rnorm = rnorm(1000))`.

Usage`ts_reg_list`**Format**

A tibble with 1000 rows and 8 variables:

x1 Random normal simulated data.

x2 Random normal simulated data.

ts_rnorm Random normal simulated data.

Index

* datasets

hsb, 4

penguins, 5

sim_longitudinal, 7

simple_ts, 6

ts_reg_list, 8

broom::tidy, 2

extract_eq, 2, 6

hsb, 4

lmer, 3

penguins, 5

print.equation, 6

sim_longitudinal, 7

simple_ts, 6

ts_reg_list, 8