Package 'ggeffects'

March 26, 2024

Type Package

Encoding UTF-8

Title Create Tidy Data Frames of Marginal Effects for 'ggplot' from Model Outputs

Version 1.5.1

Maintainer Daniel Lüdecke <d.luedecke@uke.de>

Description Compute marginal effects and adjusted predictions from statistical models and returns the result as tidy data frames. These data frames are ready to use with the 'ggplot2'-package. Effects and predictions can be calculated for many different models. Interaction terms, splines and polynomial terms are also supported. The main functions are ggpredict(), ggemmeans() and ggeffect(). There is a generic plot()-method to plot the results using 'ggplot2'.

Depends R (>= 3.6)

Imports graphics, insight (>= 0.19.8), stats, utils

Suggests AER, aod, bayestestR, betareg, brglm2, brms, broom, car, carData, clubSandwich, datawizard (>= 0.9.0), effects (>= 4.2-2), emmeans (>= 1.8.9), fixest, gam, gamlss, gamm4, gee, geepack, ggplot2, ggrepel, GLMMadaptive, glmmTMB (>= 1.1.7), gridExtra, gt, haven, httr, jsonlite, knitr, lme4 (>= 1.1-35), logistf, magrittr, margins, marginaleffects (>= 0.16.0), MASS, Matrix, mice, MCMCglmm, mgcv, nestedLogit (>= 0.3.0), nlme, nnet, ordinal, parameters, parsnip, patchwork, pscl, quantreg, rmarkdown, rms, robustbase, rstanarm, rstantools, sandwich, sdmTMB (>= 0.4.0), see, sjlabelled (>= 1.1.2), sjstats, survey, survival, testthat, tibble, tinytable (>= 0.1.0), withr, VGAM

URL https://strengejacke.github.io/ggeffects/

BugReports https://github.com/strengejacke/ggeffects/issues/

RoxygenNote 7.3.1

VignetteBuilder knitr

Config/testthat/edition 3

License MIT + file LICENSE

NeedsCompilation no

Author Daniel Lüdecke [aut, cre] (<https://orcid.org/0000-0002-8895-3206>),
Frederik Aust [ctb] (<https://orcid.org/0000-0003-4900-788X>),
Sam Crawley [ctb] (<https://orcid.org/0000-0002-7847-0411>),
Mattan S. Ben-Shachar [ctb] (<https://orcid.org/0000-0002-4287-4801>),
Sean C. Anderson [ctb] (<https://orcid.org/0000-0001-9563-1937>)

Repository CRAN

Date/Publication 2024-03-26 05:50:09 UTC

R topics documented:

`as.data.frame.ggeffects`

Adjusted predictions from regression models

Description

The **ggeffects** package computes marginal means and adjusted predicted values for the response, at the margin of specific values or levels from certain model terms. The package is built around three core functions: `predict_response()` (understanding results), `test_predictions()` (testing results for statistically significant differences) and `plot()` (communicate results).

By default, adjusted predictions or marginal means are by returned on the *response* scale, which is the easiest and most intuitive scale to interpret the results. There are other options for specific

models as well, e.g. with zero-inflation component (see documentation of the type-argument). The result is returned as consistent data frame, which is nicely printed by default. plot() can be used to easily create figures.

The main function to calculate marginal means and adjusted predictions is predict_response(). In previous versions of **ggeffects**, the functions ggpredict(), ggemmeans(), ggeffect() and ggaverage() were used to calculate marginal means and adjusted predictions. These functions are still available, but predict_response() as a "wrapper" around these functions is the preferred way to do this now.

Usage

```
## S3 method for class 'ggeffects'
as.data.frame(
  x,
  row.names = NULL,
  optional = FALSE,
  ...,
  stringsAsFactors = FALSE,
  terms_to_colnames = FALSE
)

ggaverage(
  model,
  terms,
  ci_level = 0.95,
  type = "fixed",
  typical = "mean",
  condition = NULL,
  back_transform = TRUE,
  vcov_fun = NULL,
  vcov_type = NULL,
  vcov_args = NULL,
  weights = NULL,
  verbose = TRUE,
  ...
)

ggeffect(model, terms, ci_level = 0.95, verbose = TRUE, ci.lvl = ci_level, ...)

ggemmeans(
  model,
  terms,
  ci_level = 0.95,
  type = "fixed",
  typical = "mean",
  condition = NULL,
  back_transform = TRUE,
  interval = "confidence",
```

```
  verbose = TRUE,
  ci.lvl = ci_level,
  back.transform = back_transform,
  ...
)

ggpredict(
  model,
  terms,
  ci_level = 0.95,
  type = "fixed",
  typical = "mean",
  condition = NULL,
  back_transform = TRUE,
  ppd = FALSE,
  vcov_fun = NULL,
  vcov_type = NULL,
  vcov_args = NULL,
  interval,
  verbose = TRUE,
  ci.lvl = ci_level,
  back.transform = back_transform,
  vcov.fun = vcov_fun,
  vcov.type = vcov_type,
  vcov.args = vcov_args,
  ...
)
```

Arguments

x	An object of class ggeffects, as returned by predict_response(), ggpredict(), ggeffect(), ggaverage() or ggemmeans().
row.names	NULL or a character vector giving the row names for the data frame. Missing values are not allowed.
optional	logical. If TRUE, setting row names and converting column names (to syntactic names: see make.names) is optional. Note that all of R's **base** package as.data.frame() methods use optional only for column names treatment, basically with the meaning of data.frame(*, check.names = !optional). See also the make.names argument of the matrix method.
...	Arguments are passed down to ggpredict() (further down to predict()) or ggemmeans() (and thereby to emmeans::emmeans()), If type = "simulate", ... may also be used to set the number of simulation, e.g. nsim = 500. When calling ggeffect(), further arguments passed down to effects::Effect().
stringsAsFactors	logical: should the character vector be converted to a factor?
terms_to_colnames	Logical, if TRUE, standardized column names (like "x", "group" or "facet") are replaced by the variable names of the focal predictors specified in terms.

`model` A model object, or a list of model objects.

`terms` Names of those terms from `model`, for which predictions should be displayed (so called *focal terms*). Can be:

- A character vector, specifying the names of the focal terms. This is the preferred and probably most flexible way to specify focal terms, e.g. `terms = "x [40:60]"`, to calculate predictions for the values 40 to 60.
- A list, where each element is a named vector, specifying the focal terms and their values. This is the "classical" R way to specify focal terms, e.g. `list(x = 40:60)`.
- A formula, e.g. `terms = ~ x + z`, which is internally converted to a character vector. This is probably the least flexible way, as you cannot specify representative values for the focal terms.
- A data frame representig a "data grid" or "reference grid". Predictions are then made for all combinations of the variables in the data frame.

At least one term is required to calculate effects for certain terms, maximum length is four terms, where the second to fourth term indicate the groups, i.e. predictions of first term are grouped at meaningful values or levels of the remaining terms (see `values_at()`). If `terms` is missing or `NULL`, adjusted predictions for each model term are calculated (i.e. each model term is used as single focal term). It is also possible to define specific values for focal terms, at which adjusted predictions should be calculated (see 'Details'). All remaining covariates that are not specified in `terms` are held constant (see 'Details'). See also arguments `condition` and `typical`.

`ci_level` Numeric, the level of the confidence intervals. Use `ci_level = NA` if confidence intervals should not be calculated (for instance, due to computation time). Typically, confidence intervals are based on the returned standard errors for the predictions, assuming a t- or normal distribution (based on the model and the available degrees of freedom, i.e. roughly `+/- 1.96 * SE`). See introduction of this vignette for more details.

`type` Character, indicating whether predictions should be conditioned on specific model components or not. Consequently, most options only apply for survival models, mixed effects models and/or models with zero-inflation (and their Bayesian counter-parts); only exeption is `type = "simulate"`, which is available for some other model classes as well (which respond to `simulate()`).

Note: For `brmsfit`-models with zero-inflation component, there is no `type = "zero_inflated"` nor `type = "zi_random"`; predicted values for `MixMod`-models from **GLMMadaptive** with zero-inflation component *always* condition on the zero-inflation part of the model (see 'Details').

- `"fixed"` (or `"fe"` or `"count"`)
  Predicted values are conditioned on the fixed effects or conditional model only (for mixed models: predicted values are on the population-level and *confidence intervals* are returned, i.e. `re.form = NA` when calling `predict()`). For instance, for models fitted with `zeroinfl` from **pscl**, this would return the predicted mean from the count component (without zero-inflation). For models with zero-inflation component, this type calls `predict(..., type = "link")` (however, predicted values are back-transformed to the response scale).

- `"fixed_ppd"`
  Only applies to `margin = "mean_reference"`, and only for Bayesian models of class `stanreg` or `brmsfit`. Computes the posterior predictive distribution. It is the same as setting `type = "fixed"` in combination with `ppd = TRUE`.
- `"random"` (or `"re"`)
  This only applies to mixed models, and `type = "random"` does not condition on the zero-inflation component of the model. `type = "random"` still returns population-level predictions, however, conditioned on random effects and considering individual level predictions, i.e. `re.form = NULL` when calling `predict()`. This may affect the returned predicted values, depending on whether `REML = TRUE` or `REML = FALSE` was used for model fitting. Furthermore, unlike `type = "fixed"`, intervals also consider the uncertainty in the variance parameters (the mean random effect variance, see *Johnson et al. 2014* for details) and hence can be considered as *prediction intervals*. For models with zero-inflation component, this type calls `predict(..., type = "link")` (however, predicted values are back-transformed to the response scale).
  To get predicted values for each level of the random effects groups, add the name of the related random effect term to the `terms`-argument (for more details, see this vignette).
- `"random_ppd"`
  Only applies to `margin = "mean_reference"`,, and only for Bayesian models of class `stanreg` or `brmsfit`. Computes the posterior predictive distribution. It is the same as setting `type = "random"` in combination with `ppd = TRUE`.
- `"zero_inflated"` (or `"fe.zi"` or `"zi"`)
  Predicted values are conditioned on the fixed effects and the zero-inflation component. For instance, for models fitted with `zeroinfl` from **pscl**, this would return the predicted response (`mu*(1-p)`) and for **glmmTMB**, this would return the expected value `mu*(1-p)` *without* conditioning on random effects (i.e. random effect variances are not taken into account for the confidence intervals). For models with zero-inflation component, this type calls `predict(..., type = "response")`. See 'Details'.
- `"zi_random"` (or `"re.zi"` or `"zero_inflated_random"`)
  Predicted values are conditioned on the zero-inflation component and take the random effects uncertainty into account. For models fitted with `glmmTMB()`, `hurdle()` or `zeroinfl()`, this would return the expected value `mu*(1-p)`. For **glmmTMB**, prediction intervals also consider the uncertainty in the random effects variances. This type calls `predict(..., type = "response")`. See 'Details'.
- `"zi_prob"` (or `"zi.prob"`)
  Predicted zero-inflation probability. For **glmmTMB** models with zero-inflation component, this type calls `predict(..., type = "zlink")`; models from **pscl** call `predict(..., type = "zero")` and for **GLMMadaptive**, `predict(..., type = "zero_part")` is called.
- `"simulate"` (or `"sim"`)
  Predicted values and confidence resp. prediction intervals are based on

simulations, i.e. calls to `simulate()`. This type of prediction takes all model uncertainty into account, including random effects variances. Currently supported models are objects of class `lm`, `glm`, `glmmTMB`, `wbm`, `MixMod` and `merMod`. See `...` for details on number of simulations.

- `"survival"` and `"cumulative_hazard"` (or `"surv"` and `"cumhaz"`) Applies only to `coxph`-objects from the **survial**-package and calculates the survival probability or the cumulative hazard of an event.

`typical` Character vector, naming the function to be applied to the covariates (non-focal terms) over which the effect is "averaged". The default is `"mean"`. Can be `"mean"`, `"weighted.mean"`, `"median"`, `"mode"` or `"zero"`, which call the corresponding R functions (except `"mode"`, which calls an internal function to compute the most common value); `"zero"` simply returns 0. By default, if the covariate is a factor, only `"mode"` is applicable; for all other values (including the default, `"mean"`) the reference level is returned. For character vectors, only the mode is returned. You can use a named vector to apply different functions to integer, numeric and categorical covariates, e.g. `typical = c(numeric = "median", factor = "mode")`. If `typical` is `"weighted.mean"`, weights from the model are used. If no weights are available, the function falls back to `"mean"`. **Note** that this argument is ignored for `predict_response()`, because the `margin` argument takes care of this.

`condition` Named character vector, which indicates covariates that should be held constant at specific values. Unlike `typical`, which applies a function to the covariates to determine the value that is used to hold these covariates constant, `condition` can be used to define exact values, for instance `condition = c(covariate1 = 20, covariate2 = 5)`. See 'Examples'.

`back_transform` Logical, if `TRUE` (the default), predicted values for log- or log-log transformed responses will be back-transformed to original response-scale.

`vcov_fun` Variance-covariance matrix used to compute uncertainty estimates (e.g., for confidence intervals based on robust standard errors). This argument accepts a covariance matrix, a function which returns a covariance matrix, or a string which identifies the function to be used to compute the covariance matrix.

- A (variance-covariance) matrix
- A function which returns a covariance matrix (e.g., `stats::vcov()`)
- A string which indicates the estimation type for the heteroscedasticity-consistent variance-covariance matrix, e.g. `vcov_fun = "HC0"`. Possible values are `"HC0"`, `"HC1"`, `"HC2"`, `"HC3"`, `"HC4"`, `"HC4m"`, and `"HC5"`, which will then call the `vcovHC()`-function from the **sandwich** package, using the specified type. Further possible values are `"CR0"`, `"CR1"`, `"CR1p"`, `"CR1S"`, `"CR2"`, and `"CR3"`, which will call the `vcovCR()`-function from the **clubSandwich** package.
- A string which indicates the name of the `vcov*()`-function from the **sandwich** or **clubSandwich** packages, e.g. `vcov_fun = "vcovCL"`, which is used to compute (cluster) robust standard errors for predictions.

If `NULL`, standard errors (and confidence intervals) for predictions are based on the standard errors as returned by the `predict()`-function. **Note** that probably not all model objects that work with `ggpredict()` are also supported by the **sandwich** or **clubSandwich** packages.

See details in this vignette.

vcov_type
: Character vector, specifying the estimation type for the robust covariance matrix estimation (see ?sandwich::vcovHC or ?clubSandwich::vcovCR for details). Only used when vcov_fun is a character string indicating one of the functions from those packages.

vcov_args
: List of named vectors, used as additional arguments that are passed down to vcov_fun.

weights
: Character vector, naming the weigthing variable in the data, or a vector of weights (of same length as the number of observations in the data). Only applies to margin = "empirical".

verbose
: Toggle messages or warnings.

ci.lvl, vcov.fun, vcov.type, vcov.args, back.transform
: Deprecated arguments. Please use ci_level, vcov_fun, vcov_type, vcov_args and back_transform instead.

interval
: Type of interval calculation, can either be "confidence" (default) or "prediction". May be abbreviated. Unlike *confidence intervals*, *prediction intervals* include the residual variance (sigma^2) to account for the uncertainty of predicted values. For mixed models, interval = "prediction" is the default for type = "random". When type = "fixed", the default is interval = "confidence". Note that prediction intervals are not available for all models, but only for models that work with insight::get_sigma().

ppd
: Logical, if TRUE, predictions for Stan-models are based on the posterior predictive distribution rstantools::posterior_predict(). If FALSE (the default), predictions are based on posterior draws of the linear predictor rstantools::posterior_linpred().

Details

Please see ?predict_response for details and examples.

Value

A data frame (with ggeffects class attribute) with consistent data columns:

- "x": the values of the first term in terms, used as x-position in plots.
- "predicted": the predicted values of the response, used as y-position in plots.
- "std.error": the standard error of the predictions. *Note that the standard errors are always on the link-scale, and not back-transformed for non-Gaussian models!*
- "conf.low": the lower bound of the confidence interval for the predicted values.
- "conf.high": the upper bound of the confidence interval for the predicted values.
- "group": the grouping level from the second term in terms, used as grouping-aesthetics in plots.
- "facet": the grouping level from the third term in terms, used to indicate facets in plots.

The estimated marginal means (or predicted values) are always on the response scale!

For proportional odds logistic regression (see ?MASS::polr) resp. cumulative link models (e.g., see ?ordinal::clm), an additional column "response.level" is returned, which indicates the grouping of predictions based on the level of the model's response.

Note that for convenience reasons, the columns for the intervals are always named "conf.low" and "conf.high", even though for Bayesian models credible or highest posterior density intervals are returned.

There is an as.data.frame() method for objects of class ggeffects, which has an terms_to_colnames argument, to use the term names as column names instead of the standardized names "x" etc.

---

collapse_by_group *Collapse raw data by random effect groups*

Description

This function extracts the raw data points (i.e. the data that was used to fit the model) and "averages" (i.e. "collapses") the response variable over the levels of the grouping factor given in collapse_by. Only works with mixed models.

Usage

```
collapse_by_group(grid, model, collapse_by = NULL, residuals = FALSE)
```

Arguments

grid	A data frame representing the data grid, or an object of class ggeffects, as returned by predict_response().
model	The model for which to compute partial residuals. The data grid grid should match to predictors in the model.
collapse_by	Name of the (random effects) grouping factor. Data is collapsed by the levels of this factor.
residuals	Logical, if TRUE, collapsed partial residuals instead of raw data by the levels of the grouping factor.

Value

A data frame with raw data points, averaged over the levels of the given grouping factor from the random effects. The group level of the random effect is saved in the column "random".

Examples

```
library(ggeffects)
data(efc, package = "ggeffects")
efc$e15relat <- as.factor(efc$e15relat)
efc$c161sex <- as.factor(efc$c161sex)
levels(efc$c161sex) <- c("male", "female")
model <- lme4::lmer(neg_c_7 ~ c161sex + (1 | e15relat), data = efc)
me <- predict_response(model, terms = "c161sex")
head(attributes(me)$rawdata)
collapse_by_group(me, model, "e15relat")
```

efc *Sample dataset from the EUROFAMCARE project*

Description

A SPSS sample data set, imported with the `sjlabelled::read_spss()` function.

Examples

```
# Attach EFC-data
data(efc)

# Show structure
str(efc)

# show first rows
head(efc)
```

fish *Sample data set*

Description

A sample data set, used in tests and some examples.

format.ggeffects *Print and format ggeffects-objects*

Description

A generic print-method for `ggeffects`-objects.

Usage

```
## S3 method for class 'ggeffects'
format(
  x,
  variable_labels = FALSE,
  value_labels = FALSE,
  group_name = FALSE,
  row_header_separator = ", ",
  digits = 2,
  collapse_ci = FALSE,
  collapse_tables = FALSE,
```

```
  n,
  ...
)

## S3 method for class 'ggeffects'
print(x, group_name = TRUE, digits = 2, verbose = TRUE, ...)

## S3 method for class 'ggeffects'
print_md(x, group_name = TRUE, digits = 2, ...)

## S3 method for class 'ggeffects'
print_html(
  x,
  group_name = TRUE,
  digits = 2,
  theme = NULL,
  engine = c("tt", "gt"),
  ...
)
```

Arguments

x	An object of class ggeffects, as returned by the functions from this package.
variable_labels	Logical, if TRUE variable labels are used as column headers. If FALSE, variable names are used.
value_labels	Logical, if TRUE, value labels are used as values in the table output. If FALSE, the numeric values or factor levels are used.
group_name	Logical, if TRUE, the name of further focal terms are used in the sub-headings of the table. If FALSE, only the values of the focal terms are used.
row_header_separator	Character, separator between the different subgroups in the table output.
digits	Number of digits to print.
collapse_ci	Logical, if TRUE, the columns with predicted values and confidence intervals are collapsed into one column, e.g. Predicted (95% CI).
collapse_tables	Logical, if TRUE, all tables are combined into one. The tables are not split by further focal terms, but rather are added as columns. Only works when there is more than one focal term.
n	Number of rows to print per subgroup. If NULL, a default number of rows is printed, depending on the number of subgroups.
...	Further arguments passed down to format.ggeffects(), some of them are also passed down further to insight::format_table() or insight::format_value().
verbose	Toggle messages.
theme	The theme to apply to the table. One of "grid", "striped", "bootstrap", or "darklines".

engine — The engine to use for printing. One of "tt" (default) or "gt". "tt" uses the *tinytable* package, "gt" uses the *gt* package.

Value

format() return a formatted data frame, print() prints a formatted data frame printed to the console. print_html() returns a tinytable object by default (unless changed with engine = "gt"), which is printed as HTML, markdown or LaTeX table (depending on the context from which print_html() is called, see tinytable::tt() for details).

Global Options to Customize Tables when Printing

The verbose argument can be used to display or silence messages and warnings. Furthermore, options() can be used to set defaults for the print() and print_html() method. The following options are available, which can simply be run in the console:

- ggeffects_ci_brackets: Define a character vector of length two, indicating the opening and closing parentheses that encompass the confidence intervals values, e.g. options(ggeffects_ci_brackets = c("[", "]")).
- ggeffects_collapse_ci: Logical, if TRUE, the columns with predicted values (or contrasts) and confidence intervals are collapsed into one column, e.g. options(ggeffects_collapse_ci = TRUE).
- ggeffects_collapse_p: Logical, if TRUE, the columns with predicted values (or contrasts) and p-values are collapsed into one column, e.g. options(ggeffects_collapse_p = TRUE). Note that p-values are replaced by asterisk-symbols (stars) or empty strings when ggeffects_collapse_p = TRUE, depending on the significance level.
- ggeffects_collapse_tables: Logical, if TRUE, multiple tables for subgroups are combined into one table. Only works when there is more than one focal term, e.g. options(ggeffects_collapse_tables = TRUE).
- ggeffects_output_format: String, either "text", "markdown" or "html". Defines the default output format from predict_response(). If "html", a formatted HTML table is created and printed to the view pane. "markdown" creates a markdown-formatted table inside Rmarkdown documents, and prints a text-format table to the console when used interactively. If "text" or NULL, a formatted table is printed to the console, e.g. options(ggeffects_output_format = "html").
- ggeffects_html_engine: String, either "tt" or "gt". Defines the default engine to use for printing HTML tables. If "tt", the *tinytable* package is used, if "gt", the *gt* package is used, e.g. options(ggeffects_html_engine = "gt").

Use options(<option_name> = NULL) to remove the option.

Examples

```
data(efc, package = "ggeffects")
fit <- lm(barthtot ~ c12hour + e42dep, data = efc)

# default print
predict_response(fit, "e42dep")
```

```
# surround CI values with parentheses
print(predict_response(fit, "e42dep"), ci_brackets = c("(", ")"))
# you can also use `options(ggeffects_ci_brackets = c("[", "]"))`
# to set this globally

# collapse CI columns into column with predicted values
print(predict_response(fit, "e42dep"), collapse_ci = TRUE)

# include value labels
print(predict_response(fit, "e42dep"), value_labels = TRUE)

# include variable labels in column headers
print(predict_response(fit, "e42dep"), variable_labels = TRUE)

# include value labels and variable labels
print(predict_response(fit, "e42dep"), variable_labels = TRUE, value_labels = TRUE)

data(iris)
m <- lm(Sepal.Length ~ Species * Petal.Length, data = iris)

# default print with subgroups
predict_response(m, c("Petal.Length", "Species"))

# omit name of grouping variable in subgroup table headers
print(predict_response(m, c("Petal.Length", "Species")), group_name = FALSE)

# collapse tables into one
print(predict_response(m, c("Petal.Length", "Species")), collapse_tables = TRUE, n = 3)

# increase number of digits
print(predict_response(fit, "e42dep"), digits = 5)
```

`get_title` *Get titles and labels from data*

Description

Get variable and value labels from `ggeffects`-objects. `predict_response()` saves information on variable names and value labels as additional attributes in the returned data frame. This is especially helpful for labelled data (see **sjlabelled**), since these labels can be used to set axis labels and titles.

Usage

```
get_title(x, case = NULL)

get_x_title(x, case = NULL)

get_y_title(x, case = NULL)
```

```
get_legend_title(x, case = NULL)

get_legend_labels(x, case = NULL)

get_x_labels(x, case = NULL)

get_complete_df(x, case = NULL)
```

Arguments

x	An object of class ggeffects, as returned by any ggeffects-function; for get_complete_df(), must be a list of ggeffects-objects.
case	Desired target case. Labels will automatically converted into the specified character case. See ?sjlabelled::convert_case for more details on this argument.

Value

The titles or labels as character string, or NULL, if variables had no labels; get_complete_df() returns the input list x as single data frame, where the grouping variable indicates the predicted values for each term.

Examples

```
library(ggeffects)
library(ggplot2)
data(efc, package = "ggeffects")
efc$c172code <- datawizard::to_factor(efc$c172code)
fit <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)

mydf <- predict_response(fit, terms = c("c12hour", "c161sex", "c172code"))

ggplot(mydf, aes(x = x, y = predicted, colour = group)) +
  stat_smooth(method = "lm") +
  facet_wrap(~facet, ncol = 2) +
  labs(
    x = get_x_title(mydf),
    y = get_y_title(mydf),
    colour = get_legend_title(mydf)
  )

# adjusted predictions, a list of data frames (one data frame per term)
eff <- ggeffect(fit)
eff
get_complete_df(eff)

# adjusted predictions for education only, and get x-axis-labels
mydat <- eff[["c172code"]]
ggplot(mydat, aes(x = x, y = predicted, group = group)) +
```

```
  stat_summary(fun = sum, geom = "line") +
  scale_x_discrete(labels = get_x_labels(mydat))
```

`install_latest` *Update latest ggeffects-version from R-universe (GitHub) or CRAN*

Description

This function can be used to install the latest package version of *ggeffects*, either the development version (from R-universe/GitHub) or the current version from CRAN.

Usage

```
install_latest(
  source = c("development", "cran"),
  force = FALSE,
  verbose = TRUE
)
```

Arguments

`source`	Character. Either `"development"` or `"cran"`. If `"cran"`, *ggeffects* will be installed from the default CRAN mirror returned by `getOption("repos")['CRAN']`. If `"development"` (the default), *ggeffects* is installed from the r-universe repository (https://strengejacke.r-universe.dev/).
`force`	Logical, if `FALSE`, the update will only be installed if a newer version is available. Use `force=TRUE` to force installation, even if the version number for the locally installed package is identical to the latest development-version. Only applies when `source="development"`.
`verbose`	Toggle messages.

Value

Invisible `NULL`.

Examples

```
# install latest development-version of ggeffects from the
# r-universe repository
install_latest()
```

johnson_neyman *Spotlight-analysis: Create Johnson-Neyman confidence intervals and plots*

Description

Function conduct a spotlight-analysis to create so-called Johnson-Neyman intervals. The `plot()` method can be used to visualize the results of the Johnson-Neyman test.

Usage

```
johnson_neyman(x, precision = 500, p_adjust = NULL, ...)

spotlight_analysis(x, precision = 500, p_adjust = NULL, ...)

## S3 method for class 'ggjohnson_neyman'
plot(
  x,
  colors = c("#f44336", "#2196F3"),
  show_association = TRUE,
  show_rug = FALSE,
  verbose = TRUE,
  ...
)
```

Arguments

`x`	An object of class `ggeffects`, as returned by the functions from this package.
`precision`	Number of values used for the range of the moderator variable to calculate the Johnson-Neyman interval. This argument is passed down to `pretty(..., n = precision)`. Usually, the default value of 500 is sufficient. Increasing this value will result in a smoother plot and more accurate values for the interval bounds, but can also slightly increase the computation time.
`p_adjust`	Character vector, if not `NULL`, indicates the method to adjust p-values. See `stats::p.adjust()` or `stats::p.adjust.methods` for details. Further possible adjustment methods are `"tukey"` or `"sidak"`, and for `johnson_neyman()`, `"fdr"` (or `"bh"`) and `"esarey"` (or its short-cut `"es"`) are available options. Some caution is necessary when adjusting p-value for multiple comparisons. See also section *P-value adjustment* below.
`...`	Arguments passed down to `test_predictions()` (and then probably further to `marginaleffects::slopes()`). See `?test_predictions` for further details.
`colors`	Colors used for the plot. Must be a vector with two color values. Only used if `show_association = TRUE`.
`show_association`	Logical, if `TRUE`, highlights the range where values of the moderator are positively or negtatively associated with the outcome.

show_rug	Logical, if TRUE, adds a rug with raw data of the moderator variable to the plot. This helps visualizing its distribution.
verbose	Show/hide printed message for plots.

Details

The Johnson-Neyman intervals help to understand where slopes are significant in the context of interactions in regression models. Thus, the interval is only useful if the model contains at least one interaction term. The function accepts the results of a call to `predict_response()`. The *first* and the *last* focal term used in the `terms` argument of `predict_response()` must be numeric. The function will then test the slopes of the first focal terms against zero, for different moderator values of the last focal term. If only one numeric focal term is given, the function will create contrasts by levels of the categorical focal term. Use `plot()` to create a plot of the results.

To avoid misleading interpretations of the plot, we speak of "positive" and "negative" associations, respectively, and "no clear" associations (instead of "significant" or "non-significant"). This should prevent the user from considering a non-significant range of values of the moderator as "accepting the null hypothesis".

Value

A data frame including contrasts of the `test_predictions()` for the given interaction terms; for `plot()`, returns a Johnson-Neyman plot.

P-value adjustment for multiple comparisons

Note that p-value adjustment for methods supported by `p.adjust()` (see also `p.adjust.methods`), each row is considered as one set of comparisons, no matter which `test` was specified. That is, for instance, when `test_predictions()` returns eight rows of predictions (when `test = NULL`), and `p_adjust = "bonferroni"`, the p-values are adjusted in the same way as if we had a test of pairwise comparisons (`test = "pairwise"`) where eight rows of comparisons are returned. For methods `"tukey"` or `"sidak"`, a rank adjustment is done based on the number of combinations of levels from the focal predictors in `terms`. Thus, the latter two methods may be useful for certain tests only, in particular pairwise comparisons.

For `johnson_neyman()`, the only available adjustment methods are `"fdr"` (or `"bh"`) (*Benjamini & Hochberg (1995)*) and `"esarey"` (or `"es"`) (*Esarey and Sumner 2017*). These usually return similar results. The major difference is that `"fdr"` can be slightly faster and more stable in edge cases, however, confidence intervals are not updated. Only the p-values are adjusted. `"esarey"` is slower, but confidence intervals are updated as well.

References

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. Multivariate Behavioral Research, 40(3), 373-400. doi: 10.1207/s15327906mbr4003_5

Esarey, J., & Sumner, J. L. (2017). Marginal effects in interaction models: Determining and controlling the false positive rate. Comparative Political Studies, 1–33. Advance online publication. doi: 10.1177/0010414017730080

Johnson, P.O. & Fay, L.C. (1950). The Johnson-Neyman technique, its theory and application. Psychometrika, 15, 349-367. doi: 10.1007/BF02288864

McCabe CJ, Kim DS, King KM. Improving Present Practices in the Visual Display of Interactions. Advances in Methods and Practices in Psychological Science. 2018;1(2):147-165. doi:10.1177/2515245917746792

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., & McClelland, G. H. (2013). Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression. Journal of Marketing Research, 50(2), 277–288. doi:10.1509/jmr.12.0420

Examples

```
## Not run:
data(efc, package = "ggeffects")
efc$c172code <- as.factor(efc$c172code)
m <- lm(neg_c_7 ~ c12hour * barthtot * c172code, data = efc)

pr <- predict_response(m, c("c12hour", "barthtot"))
johnson_neyman(pr)
plot(johnson_neyman(pr))

pr <- predict_response(m, c("c12hour", "c172code", "barthtot"))
johnson_neyman(pr)
plot(johnson_neyman(pr))

# robust standard errors
if (requireNamespace("sandwich")) {
  johnson_neyman(pr, vcov = sandwich::vcovHC)
}

## End(Not run)
```

---

lung2 *Sample data set*

Description

A sample data set, used in tests and examples for survival models. This dataset is originally included in the **survival** package, but for convenience reasons it is also available in this package.

---

new_data *Create a data frame from all combinations of predictor values*

Description

Create a data frame for the "newdata"-argument that contains all combinations of values from the terms in questions. Similar to `expand.grid()`. The `terms`-argument accepts all shortcuts for representative values as in `predict_response()`.

Usage

```
new_data(model, terms, typical = "mean", condition = NULL, ...)

data_grid(model, terms, typical = "mean", condition = NULL, ...)
```

Arguments

model	A fitted model object.
terms	Character vector with the names of those terms from model for which all combinations of values should be created. This argument works in the same way as the terms argument in predict_response(). See also this vignette.
typical	Character vector, naming the function to be applied to the covariates (non-focal terms) over which the effect is "averaged". The default is "mean". Can be "mean", "weighted.mean", "median", "mode" or "zero", which call the corresponding R functions (except "mode", which calls an internal function to compute the most common value); "zero" simply returns 0. By default, if the covariate is a factor, only "mode" is applicable; for all other values (including the default, "mean") the reference level is returned. For character vectors, only the mode is returned. You can use a named vector to apply different functions to integer, numeric and categorical covariates, e.g. typical = c(numeric = "median", factor = "mode"). If typical is "weighted.mean", weights from the model are used. If no weights are available, the function falls back to "mean". **Note** that this argument is ignored for predict_response(), because the margin argument takes care of this.
condition	Named character vector, which indicates covariates that should be held constant at specific values. Unlike typical, which applies a function to the covariates to determine the value that is used to hold these covariates constant, condition can be used to define exact values, for instance condition = c(covariate1 = 20, covariate2 = 5). See 'Examples'.
...	Currently not used.

Value

A data frame containing one row for each combination of values of the supplied variables.

Examples

```
data(efc, package = "ggeffects")
fit <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)
new_data(fit, c("c12hour [meansd]", "c161sex"))

nd <- new_data(fit, c("c12hour [meansd]", "c161sex"))
pr <- predict(fit, type = "response", newdata = nd)
nd$predicted <- pr
nd

# compare to
```

```
predict_response(fit, c("c12hour [meansd]", "c161sex"))
```

---

plot *Plot ggeffects-objects*

Description

A generic plot-method for ggeffects-objects.

Usage

```
## S3 method for class 'ggeffects'
plot(
  x,
  show_ci = TRUE,
  ci_style = c("ribbon", "errorbar", "dash", "dot"),
  show_data = FALSE,
  show_residuals = FALSE,
  show_residuals_line = FALSE,
  data_labels = FALSE,
  limit_range = FALSE,
  collapse_group = FALSE,
  show_legend = TRUE,
  show_title = TRUE,
  show_x_title = TRUE,
  show_y_title = TRUE,
  case = NULL,
  colors = NULL,
  alpha = 0.15,
  dot_alpha = 0.35,
  jitter = NULL,
  dodge = 0.25,
  dot_size = NULL,
  line_size = NULL,
  use_theme = TRUE,
  log_y = FALSE,
  connect_lines = FALSE,
  facets,
  grid,
  one_plot = TRUE,
  verbose = TRUE,
  ci = show_ci,
  ci.style = ci_style,
  rawdata = show_data,
  add.data = show_data,
  residuals = show_residuals,
```

```
  residuals.line = show_residuals_line,
  label.data = data_labels,
  limit.range = limit_range,
  collapse.group = collapse_group,
  dot.alpha = dot_alpha,
  dot.size = dot_size,
  line.size = line_size,
  connect.lines = connect_lines,
  show.title = show_title,
  show.x.title = show_x_title,
  show.y.title = show_y_title,
  use.theme = use_theme,
  show.legend = show_legend,
  one.plot = one_plot,
  log.y = log_y,
  ...
)

theme_ggeffects(base_size = 11, base_family = "")

show_pals()
```

Arguments

x	An object of class ggeffects, as returned by the functions from this package.
show_ci	Logical, if TRUE, confidence bands (for continuous variables at x-axis) resp. error bars (for factors at x-axis) are plotted.
ci_style	Character vector, indicating the style of the confidence bands. May be either "ribbon", "errorbar", "dash" or "dot", to plot a ribbon, error bars, or dashed or dotted lines as confidence bands.
show_data	Logical, if TRUE, a layer with raw data from response by predictor on the x-axis, plotted as point-geoms, is added to the plot.
show_residuals	Logical, if TRUE, a layer with partial residuals is added to the plot. See vignette Effect Displays with Partial Residuals. from **effects** for more details on partial residual plots.
show_residuals_line	Logical, if TRUE, a loess-fit line is added to the partial residuals plot. Only applies if residuals is TRUE.
data_labels	Logical, if TRUE and row names in data are available, data points will be labelled by their related row name.
limit_range	Logical, if TRUE, limits the range of the prediction bands to the range of the data.
collapse_group	For mixed effects models, name of the grouping variable of random effects. If collapse_group = TRUE, data points "collapsed" by the first random effect groups are added to the plot. Else, if collapse_group is a name of a group factor, data is collapsed by that specific random effect. See collapse_by_group() for further details.

show_legend — Logical, shows or hides the plot legend.

show_title — Logical, shows or hides the plot title-

show_x_title — Logical, shows or hides the plot title for the x-axis.

show_y_title — Logical, shows or hides the plot title for the y-axis.

case — Desired target case. Labels will automatically converted into the specified character case. See ?sjlabelled::convert_case for more details on this argument.

colors — Character vector with color values in hex-format, valid color value names (see demo("colors")) or a name of a ggeffects-color-palette.

Following options are valid for colors:

- If not specified, the color brewer palette "Set1" will be used.
- If "gs", a greyscale will be used.
- If "bw", the plot is black/white and uses different line types to distinguish groups.
- There are some pre-defined color-palettes in this package that can be used, e.g. colors = "metro". See show_pals() to show all available palettes.
- Else specify own color values or names as vector (e.g. colors = c("#f00000", "#00ff00")).

alpha — Alpha value for the confidence bands.

dot_alpha — Alpha value for data points, when show_data = TRUE.

jitter — Numeric, between 0 and 1. If not NULL and show_data = TRUE, adds a small amount of random variation to the location of data points dots, to avoid overplotting. Hence the points don't reflect exact values in the data. May also be a numeric vector of length two, to add different horizontal and vertical jittering. For binary outcomes, raw data is not jittered by default to avoid that data points exceed the axis limits.

dodge — Value for offsetting or shifting error bars, to avoid overlapping. Only applies, if a factor is plotted at the x-axis (in such cases, the confidence bands are replaced by error bars automatically), or if ci_style = "errorbars".

dot_size — Numeric, size of the point geoms.

line_size — Numeric, size of the line geoms.

use_theme — Logical, if TRUE, a slightly tweaked version of ggplot's minimal-theme, theme_ggeffects(), is applied to the plot. If FALSE, no theme-modifications are applied.

log_y — Logical, if TRUE, the y-axis scale is log-transformed. This might be useful for binomial models with predicted probabilities on the y-axis.

connect_lines — Logical, if TRUE and plot has point-geoms with error bars (this is usually the case when the x-axis is discrete), points of same groups will be connected with a line.

facets, grid — Logical, defaults to TRUE if x has a column named facet, and defaults to FALSE if x has no such column. Set facets = TRUE to wrap the plot into facets even for grouping variables (see 'Examples'). grid is an alias for facets.

one_plot — Logical, if TRUE and x has a panel column (i.e. when four terms were used), a single, integrated plot is produced.

verbose — Logical, toggle warnings and messages.

ci, add.data, rawdata, residuals, residuals.line, label.data, limit.range, collapse.group, dot.alpha, do — Deprecated arguments. Use show_ci, show_data, show_residuals, show_residuals_line, data_labels, limit_range, collapse_group, dot_alpha, dot_size, line_size, connect_lines, show_title, show_x_title, show_y_title, use_theme, ci_style, show_legend, log_y and one_plot instead.

... — Further arguments passed down to ggplot::scale_y*(), to control the appearance of the y-axis.

base_size — Base font size.

base_family — Base font family.

Details

For proportional odds logistic regression (see ?MASS::polr) or cumulative link models in general, plots are automatically facetted by response.level, which indicates the grouping of predictions based on the level of the model's response.

Value

A ggplot2-object.

Partial Residuals

For **generalized linear models** (glms), residualized scores are computed as inv.link(link(Y) + r) where Y are the predicted values on the response scale, and r are the *working* residuals.

For (generalized) linear **mixed models**, the random effect are also partialled out.

Note

Load library(ggplot2) and use theme_set(theme_ggeffects()) to set the **ggeffects**-theme as default plotting theme. You can then use further plot-modifiers, e.g. from **sjPlot**, like legend_style() or font_size() without losing the theme-modifications.

There are pre-defined colour palettes in this package. Use show_pals() to show all available colour palettes.

Examples

```
library(sjlabelled)
data(efc)
efc$c172code <- as_label(efc$c172code)
fit <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)

dat <- predict_response(fit, terms = "c12hour")
plot(dat)


# facet by group, use pre-defined color palette
```

```
dat <- predict_response(fit, terms = c("c12hour", "c172code"))
plot(dat, facet = TRUE, colors = "hero")

# don't use facets, b/w figure, w/o confidence bands
dat <- predict_response(fit, terms = c("c12hour", "c172code"))
plot(dat, colors = "bw", show_ci = FALSE)

# factor at x axis, plot exact data points and error bars
dat <- predict_response(fit, terms = c("c172code", "c161sex"))
plot(dat)

# for three variables, automatic facetting
dat <- predict_response(fit, terms = c("c12hour", "c172code", "c161sex"))
plot(dat)


# show all color palettes
show_pals()
```

`pool_comparisons` *Pool contrasts and comparisons from* `test_predictions()`

Description

This function "pools" (i.e. combines) multiple ggcomparisons objects, returned by `test_predictions()`, in a similar fashion as `mice::pool()`.

Usage

```
pool_comparisons(x, ...)
```

Arguments

x	A list of ggcomparisons objects, as returned by `test_predictions()`.
...	Currently not used.

Details

Averaging of parameters follows Rubin's rules (*Rubin, 1987, p. 76*).

Value

A data frame with pooled comparisons or contrasts of predictions.

References

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

Examples

```
data("nhanes2", package = "mice")
imp <- mice::mice(nhanes2, printFlag = FALSE)
comparisons <- lapply(1:5, function(i) {
  m <- lm(bmi ~ age + hyp + chl, data = mice::complete(imp, action = i))
  test_predictions(m, "age")
})
pool_comparisons(comparisons)
```

`pool_predictions` *Pool Predictions or Estimated Marginal Means*

Description

This function "pools" (i.e. combines) multiple `ggeffects` objects, in a similar fashion as `mice::pool()`.

Usage

```
pool_predictions(x, ...)
```

Arguments

`x`	A list of `ggeffects` objects, as returned by `predict_response()`.
`...`	Currently not used.

Details

Averaging of parameters follows Rubin's rules (*Rubin, 1987, p. 76*).

Value

A data frame with pooled predictions.

References

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons.

Examples

```
# example for multiple imputed datasets
data("nhanes2", package = "mice")
imp <- mice::mice(nhanes2, printFlag = FALSE)
predictions <- lapply(1:5, function(i) {
  m <- lm(bmi ~ age + hyp + chl, data = mice::complete(imp, action = i))
```

```
  predict_response(m, "age")
})
pool_predictions(predictions)
```

`predict_response` *Adjusted predictions and estimated marginal means from regression models*

Description

The **ggeffects** package computes marginal means and adjusted predicted values for the response, at the margin of specific values or levels from certain model terms. The package is built around three core functions: `predict_response()` (understanding results), `test_predictions()` (importance of results) and `plot()` (communicate results).

By default, adjusted predictions or marginal means are returned on the *response* scale, which is the easiest and most intuitive scale to interpret the results. There are other options for specific models as well, e.g. with zero-inflation component (see documentation of the `type`-argument). The result is returned as structured data frame, which is nicely printed by default. `plot()` can be used to easily create figures.

The main function to calculate marginal means and adjusted predictions is `predict_response()`, which returns adjusted predictions, marginal means or averaged counterfactual predictions depending on value of the `margin`-argument.

In previous versions of **ggeffects**, the functions `ggpredict()`, `ggemmeans()`, `ggeffect()` and `ggaverage()` were used to calculate marginal means and adjusted predictions. These functions are still available, but `predict_response()` as a "wrapper" around these functions is the preferred way to calculate marginal means and adjusted predictions now.

Usage

```
predict_response(
  model,
  terms,
  margin = "mean_reference",
  ci_level = 0.95,
  type = "fixed",
  condition = NULL,
  back_transform = TRUE,
  ppd = FALSE,
  vcov_fun = NULL,
  vcov_type = NULL,
  vcov_args = NULL,
  weights = NULL,
  interval,
  verbose = TRUE,
  ...
)
```

Arguments

model	A model object.
terms	Names of those terms from model, for which predictions should be displayed (so called *focal terms*). Can be: • A character vector, specifying the names of the focal terms. This is the preferred and probably most flexible way to specify focal terms, e.g. terms = "x [40:60]", to calculate predictions for the values 40 to 60. • A list, where each element is a named vector, specifying the focal terms and their values. This is the "classical" R way to specify focal terms, e.g. list(x = 40:60). • A formula, e.g. terms = ~ x + z, which is internally converted to a character vector. This is probably the least flexible way, as you cannot specify representative values for the focal terms. • A data frame representig a "data grid" or "reference grid". Predictions are then made for all combinations of the variables in the data frame. term at least requires one variable name. The maximum length is four terms, where the second to fourth term indicate the groups, i.e. predictions of first term are grouped at meaningful values or levels of the remaining terms (see values_at()). It is also possible to define specific values for focal terms, at which adjusted predictions should be calculated (see details below). All remaining covariates that are not specified in terms are "marginalized", see the margin argument. See also argument condition to fix non-focal terms to specific values.
margin	Character string, indicating how to marginalize over the *non-focal* predictors, i.e. those variables that are *not* specified in terms. Possible values are "mean_reference", "mean_mode", "marginalmeans" and "empirical" (or "counterfactual", aka average "counterfactual" predictions). You can set a default-option for the margin argument via options(), e.g. options(ggeffects_margin = "empirical"), so you don't have to specify your preferred marginalization method each time you call predict_response(). See details in the documentation below.
ci_level	Numeric, the level of the confidence intervals. Use ci_level = NA if confidence intervals should not be calculated (for instance, due to computation time). Typically, confidence intervals are based on the returned standard errors for the predictions, assuming a t- or normal distribution (based on the model and the available degrees of freedom, i.e. roughly +/- 1.96 * SE). See introduction of this vignette for more details.
type	Character, indicating whether predictions should be conditioned on specific model components or not. Consequently, most options only apply for survival models, mixed effects models and/or models with zero-inflation (and their Bayesian counter-parts); only exeption is type = "simulate", which is available for some other model classes as well (which respond to simulate()). **Note 1:** For brmsfit-models with zero-inflation component, there is no type = "zero_inflated" nor type = "zi_random"; predicted values for MixMod-models from **GLMMadaptive** with zero-inflation component *always* condition on the zero-inflation part of the model (see 'Details').

Note 2: If `margin = "empirical"` (i.e. counterfactual predictions), the `type` argument is handled differently. It is set to `"response"` by default, and usually accepts all values from the `type`-argument of the model's respective `predict()` method. E.g., passing a `glm` object would allow the options `"response"`, `"link"`, and `"terms"`. Thus, the following options apply to `predict_response()` when `margin` is *not* `"empirical"`, and are passed to `ggpredict()` or `ggemmeans()`, respectively (depending on the value of `margin`):

- `"fixed"` (or `"fe"` or `"count"`)
  Predicted values are conditioned on the fixed effects or conditional model only (for mixed models: predicted values are on the population-level and *confidence intervals* are returned, i.e. `re.form = NA` when calling `predict()`). For instance, for models fitted with `zeroinfl` from **pscl**, this would return the predicted mean from the count component (without zero-inflation). For models with zero-inflation component, this type calls `predict(..., type = "link")` (however, predicted values are back-transformed to the response scale).
- `"fixed_ppd"`
  Only applies to `margin = "mean_reference"`, and only for Bayesian models of class `stanreg` or `brmsfit`. Computes the posterior predictive distribution. It is the same as setting `type = "fixed"` in combination with `ppd = TRUE`.
- `"random"` (or `"re"`)
  This only applies to mixed models, and `type = "random"` does not condition on the zero-inflation component of the model. `type = "random"` still returns population-level predictions, however, conditioned on random effects and considering individual level predictions, i.e. `re.form = NULL` when calling `predict()`. This may affect the returned predicted values, depending on whether `REML = TRUE` or `REML = FALSE` was used for model fitting. Furthermore, unlike `type = "fixed"`, intervals also consider the uncertainty in the variance parameters (the mean random effect variance, see *Johnson et al. 2014* for details) and hence can be considered as *prediction intervals*. For models with zero-inflation component, this type calls `predict(..., type = "link")` (however, predicted values are back-transformed to the response scale).
  To get predicted values for each level of the random effects groups, add the name of the related random effect term to the `terms`-argument (for more details, see this vignette).
- `"random_ppd"`
  Only applies to `margin = "mean_reference"`,, and only for Bayesian models of class `stanreg` or `brmsfit`. Computes the posterior predictive distribution. It is the same as setting `type = "random"` in combination with `ppd = TRUE`.
- `"zero_inflated"` (or `"fe.zi"` or `"zi"`)
  Predicted values are conditioned on the fixed effects and the zero-inflation component. For instance, for models fitted with `zeroinfl` from **pscl**, this would return the predicted response (`mu*(1-p)`) and for **glmmTMB**, this would return the expected value `mu*(1-p)` *without* conditioning on random effects (i.e. random effect variances are not taken into account for the confi-

dence intervals). For models with zero-inflation component, this type calls `predict(..., type = "response")`. See 'Details'.

- `"zi_random"` (or `"re.zi"` or `"zero_inflated_random"`)
  Predicted values are conditioned on the zero-inflation component and take the random effects uncertainty into account. For models fitted with `glmmTMB()`, `hurdle()` or `zeroinfl()`, this would return the expected value `mu*(1-p)`. For **glmmTMB**, prediction intervals also consider the uncertainty in the random effects variances. This type calls `predict(..., type = "response")`. See 'Details'.
- `"zi_prob"` (or `"zi.prob"`)
  Predicted zero-inflation probability. For **glmmTMB** models with zero-inflation component, this type calls `predict(..., type = "zlink")`; models from **pscl** call `predict(..., type = "zero")` and for **GLMMadaptive**, `predict(..., type = "zero_part")` is called.
- `"simulate"` (or `"sim"`)
  Predicted values and confidence resp. prediction intervals are based on simulations, i.e. calls to `simulate()`. This type of prediction takes all model uncertainty into account, including random effects variances. Currently supported models are objects of class `lm`, `glm`, `glmmTMB`, `wbm`, `MixMod` and `merMod`. See `...` for details on number of simulations.
- `"survival"` and `"cumulative_hazard"` (or `"surv"` and `"cumhaz"`)
  Applies only to `coxph`-objects from the **survial**-package and calculates the survival probability or the cumulative hazard of an event.

When `margin = "empirical"`, the `type` argument accepts all values from the `type`-argument of the model's respective `predict()`-method.

`condition` Named character vector, which indicates covariates that should be held constant at specific values. Unlike `typical`, which applies a function to the covariates to determine the value that is used to hold these covariates constant, `condition` can be used to define exact values, for instance `condition = c(covariate1 = 20, covariate2 = 5)`. See 'Examples'.

`back_transform` Logical, if `TRUE` (the default), predicted values for log- or log-log transformed responses will be back-transformed to original response-scale.

`ppd` Logical, if `TRUE`, predictions for Stan-models are based on the posterior predictive distribution `rstantools::posterior_predict()`. If `FALSE` (the default), predictions are based on posterior draws of the linear predictor `rstantools::posterior_linpred()`.

`vcov_fun` Variance-covariance matrix used to compute uncertainty estimates (e.g., for confidence intervals based on robust standard errors). This argument accepts a covariance matrix, a function which returns a covariance matrix, or a string which identifies the function to be used to compute the covariance matrix.

- A (variance-covariance) matrix
- A function which returns a covariance matrix (e.g., `stats::vcov()`)
- A string which indicates the estimation type for the heteroscedasticity-consistent variance-covariance matrix, e.g. `vcov_fun = "HC0"`. Possible values are `"HC0"`, `"HC1"`, `"HC2"`, `"HC3"`, `"HC4"`, `"HC4m"`, and `"HC5"`, which will then call the `vcovHC()`-function from the **sandwich** package, using the specified type. Further possible values are `"CR0"`, `"CR1"`, `"CR1p"`, `"CR1S"`,

"CR2", and "CR3", which will call the vcovCR()-function from the **clubSandwich** package.

- A string which indicates the name of the vcov*()-function from the **sandwich** or **clubSandwich** packages, e.g. vcov_fun = "vcovCL", which is used to compute (cluster) robust standard errors for predictions.

If NULL, standard errors (and confidence intervals) for predictions are based on the standard errors as returned by the predict()-function. **Note** that probably not all model objects that work with ggpredict() are also supported by the **sandwich** or **clubSandwich** packages.

See details in this vignette.

vcov_type — Character vector, specifying the estimation type for the robust covariance matrix estimation (see ?sandwich::vcovHC or ?clubSandwich::vcovCR for details). Only used when vcov_fun is a character string indicating one of the functions from those packages.

vcov_args — List of named vectors, used as additional arguments that are passed down to vcov_fun.

weights — Character vector, naming the weigthing variable in the data, or a vector of weights (of same length as the number of observations in the data). Only applies to margin = "empirical".

interval — Type of interval calculation, can either be "confidence" (default) or "prediction". May be abbreviated. Unlike *confidence intervals*, *prediction intervals* include the residual variance (sigma^2) to account for the uncertainty of predicted values. For mixed models, interval = "prediction" is the default for type = "random". When type = "fixed", the default is interval = "confidence". Note that prediction intervals are not available for all models, but only for models that work with insight::get_sigma().

verbose — Toggle messages or warnings.

... — If margin is set to "mean_reference" or "mean_mode", arguments are passed down to ggpredict() (further down to predict()); for margin = "marginalmeans", further arguments passed down to ggemmeans() and thereby to emmeans::emmeans(); if margin = "empirical", further arguments are passed down to marginaleffects::avg_predictions(If type = "simulate", ... may also be used to set the number of simulation, e.g. nsim = 500. When calling ggeffect(), further arguments passed down to effects::Effect().

Value

A data frame (with ggeffects class attribute) with consistent data columns:

- "x": the values of the first term in terms, used as x-position in plots.
- "predicted": the predicted values of the response, used as y-position in plots.
- "std.error": the standard error of the predictions. *Note that the standard errors are always on the link-scale, and not back-transformed for non-Gaussian models!*
- "conf.low": the lower bound of the confidence interval for the predicted values.
- "conf.high": the upper bound of the confidence interval for the predicted values.

- "group": the grouping level from the second term in terms, used as grouping-aesthetics in plots.
- "facet": the grouping level from the third term in terms, used to indicate facets in plots.

The estimated marginal means (or predicted values) are always on the response scale!

For proportional odds logistic regression (see ?MASS::polr) resp. cumulative link models (e.g., see ?ordinal::clm), an additional column "response.level" is returned, which indicates the grouping of predictions based on the level of the model's response.

Note that for convenience reasons, the columns for the intervals are always named "conf.low" and "conf.high", even though for Bayesian models credible or highest posterior density intervals are returned.

There is an as.data.frame() method for objects of class ggeffects, which has an terms_to_colnames argument, to use the term names as column names instead of the standardized names "x" etc.

Supported Models

A list of supported models can be found at the package website. Support for models varies by marginalization method (the margin argument), i.e. although predict_response() supports most models, some models are only supported exclusively by one of the four downstream functions (ggpredict(), ggemmeans(), ggeffect() or ggaverage()). This means that not all models work for every margin option of predict_response().

Holding covariates at constant values, or how to marginalize over the *non-focal* predictors

predict_response() is a wrapper around ggpredict(), ggemmeans() and ggaverage(). Depending on the value of the margin argument, predict_response() calls one of those functions. The margin argument indicates how to marginalize over the *non-focal* predictors, i.e. those variables that are *not* specified in terms. Possible values are:

- "mean_reference" and "mean_mode": For "mean_reference", non-focal predictors are set to their mean (numeric variables), reference level (factors), or "most common" value (mode) in case of character vectors. For "mean_mode", non-focal predictors are set to their mean (numeric variables) or mode (factors, or "most common" value in case of character vectors).

  These predictons represent a rather "theoretical" view on your data, which does not necessarily exactly reflect the characteristics of your sample. It helps answer the question, "What is the predicted value of the response at meaningful values or levels of my focal terms for a 'typical' observation in my data?", where 'typical' refers to certain characteristics of the remaining predictors.
- "marginalmeans": non-focal predictors are set to their mean (numeric variables) or averaged over the levels or "values" for factors and character vectors. Averaging over the factor levels of non-focal terms computes a kind of "weighted average" for the values at which these terms are hold constant. Thus, non-focal categorical terms are conditioned on "weighted averages" of their levels.

  These predictions come closer to the sample, because all possible values and levels of the non-focal predictors are taken into account. It would answer the question, "What is the predicted value of the response at meaningful values or levels of my focal terms for an 'average' observation in my data?". It refers to randomly picking a subject of your sample and the result you get on average.

- `"empirical"` (or `"counterfactual"`): non-focal predictors are averaged over the observations in the sample. The response is predicted for each subject in the data and predicted values are then averaged across all subjects, aggregated/grouped by the focal terms. In particular, averaging is applied to *counterfactual predictions* (Dickerman and Hernan 2020). There is a more detailed description in this vignette.

  Counterfactual predictions are useful, insofar as the results can also be transferred to other contexts. It answers the question, "What is the predicted value of the response at meaningful values or levels of my focal terms for the 'average' observation in the population?". It does not only refer to the actual data in your sample, but also "what would be if" we had more data, or if we had data from a different population. This is where "counterfactual" refers to.

You can set a default-option for the `margin` argument via `options()`, e.g. `options(ggeffects_margin = "empirical")`, so you don't have to specify your "default" marginalization method each time you call `predict_response()`. Use `options(ggeffects_margin = NULL)` to remove that setting.

The `condition` argument can be used to fix non-focal terms to specific values.

Marginal Means and Adjusted Predictions at Specific Values

Meaningful values of focal terms can be specified via the `terms` argument. Specifying meaningful or representative values as string pattern is the preferred way in the **ggeffects** package. However, it is also possible to use a `list()` for the focal terms if prefer the "classical" R way, which is described in this vignette.

Indicating levels in square brackets allows for selecting only certain groups or values resp. value ranges. The term name and the start of the levels in brackets must be separated by a whitespace character, e.g. `terms = c("age", "education [1,3]")`. Numeric ranges, separated with colon, are also allowed: `terms = c("education", "age [30:60]")`. The stepsize for ranges can be adjusted using `by`, e.g. `terms = "age [30:60 by=5]"`.

The `terms` argument also supports the same shortcuts as the `values` argument in `values_at()`. So `terms = "age [meansd]"` would return predictions for the values one standard deviation below the mean age, the mean age and one SD above the mean age. `terms = "age [quart2]"` would calculate predictions at the value of the lower, median and upper quartile of age.

Furthermore, it is possible to specify a function name. Values for predictions will then be transformed, e.g. `terms = "income [exp]"`. This is useful when model predictors were transformed for fitting the model and should be back-transformed to the original scale for predictions. It is also possible to define own functions (see this vignette).

Instead of a function, it is also possible to define the name of a variable with specific values, e.g. to define a vector `v = c(1000, 2000, 3000)` and then use `terms = "income [v]"`.

You can take a random sample of any size with `sample=n`, e.g `terms = "income [sample=8]"`, which will sample eight values from all possible values of the variable `income`. This option is especially useful for plotting predictions at certain levels of random effects group levels, where the group factor has many levels that can be completely plotted. For more details, see this vignette.

Finally, numeric vectors for which no specific values are given, a "pretty range" is calculated (see `pretty_range()`), to avoid memory allocation problems for vectors with many unique values. If a numeric vector is specified as second or third term (i.e. if this vector represents a grouping structure), representative values (see `values_at()`) are chosen (unless other values are specified). If all values for a numeric vector should be used to compute predictions, you may use e.g. `terms = "age [all]"`. See also package vignettes.

To create a pretty range that should be smaller or larger than the default range (i.e. if no specific values would be given), use the n tag, e.g. terms="age [n=5]" or terms="age [n=12]". Larger values for n return a larger range of predicted values.

Bayesian Regression Models

predict_response() also works with **Stan**-models from the **rstanarm** or **brms**-packages. The predicted values are the median value of all drawn posterior samples. The confidence intervals for Stan-models are Bayesian predictive intervals. By default (i.e. ppd = FALSE), the predictions are based on rstantools::posterior_linpred() and hence have some limitations: the uncertainty of the error term is not taken into account. The recommendation is to use the posterior predictive distribution (rstantools::posterior_predict()).

Zero-Inflated and Zero-Inflated Mixed Models with brms

Models of class brmsfit always condition on the zero-inflation component, if the model has such a component. Hence, there is no type = "zero_inflated" nor type = "zi_random" for brmsfit-models, because predictions are based on draws of the posterior distribution, which already account for the zero-inflation part of the model.

Zero-Inflated and Zero-Inflated Mixed Models with glmmTMB

If model is of class glmmTMB, hurdle, zeroinfl or zerotrunc, and margin is *not* set to "empirical", simulations from a multivariate normal distribution (see ?MASS::mvrnorm) are drawn to calculate mu*(1-p). Confidence intervals are then based on quantiles of these results. For type = "zi_random", prediction intervals also take the uncertainty in the random-effect paramters into account (see also *Brooks et al. 2017*, pp.391-392 for details).

An alternative for models fitted with **glmmTMB** that take all model uncertainties into account are simulations based on simulate(), which is used when type = "simulate" (see *Brooks et al. 2017*, pp.392-393 for details).

Finally, if margin = "empirical", the returned predictions are already conditioned on the zero-inflation part (and possible random effects) of the model, thus these are most comparable to the type = "simulate" option. In other words, if all model components should be taken into account for predictions, you should consider using margin = "empirical".

MixMod-models from GLMMadaptive

Predicted values for the fixed effects component (type = "fixed" or type = "zero_inflated") are based on predict(..., type = "mean_subject"), while predicted values for random effects components (type = "random" or type = "zi_random") are calculated with predict(..., type = "subject_specific") (see ?GLMMadaptive::predict.MixMod for details). The latter option requires the response variable to be defined in the newdata-argument of predict(), which will be set to its typical value (see values_at()).

Multinomial Models

polr, clm models, or more generally speaking, models with ordinal or multinominal outcomes, have an additional column response.level, which indicates with which level of the response variable the predicted values are associated.

Note

Printing Results

The `print()` method gives a clean output (especially for predictions by groups), and indicates at which values covariates were held constant. Furthermore, the `print()` method has several arguments to customize the output. See this vignette for details.

Limitations

The support for some models, for example from package **MCMCglmm**, is rather experimental and may fail for certain models. If you encounter any errors, please file an issue at Github.

References

- Brooks ME, Kristensen K, Benthem KJ van, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal. 2017;9: 378-400.
- Johnson PC, O'Hara RB. 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. Methods Ecol Evol, 5: 944-946.
- Dickerman BA, Hernan, MA. Counterfactual prediction is not only for causal inference. Eur J Epidemiol 35, 615–617 (2020).

Examples

```
library(sjlabelled)
data(efc)
fit <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)

predict_response(fit, terms = "c12hour")
predict_response(fit, terms = c("c12hour", "c172code"))
# more compact table layout for printing
out <- predict_response(fit, terms = c("c12hour", "c172code", "c161sex"))
print(out, collapse_table = TRUE)

# specified as formula
predict_response(fit, terms = ~ c12hour + c172code + c161sex)

# only range of 40 to 60 for variable 'c12hour'
predict_response(fit, terms = "c12hour [40:60]")

# terms as named list
predict_response(fit, terms = list(c12hour = 40:60))

# covariate "neg_c_7" is held constant at a value of 11.84 (its mean value).
# To use a different value, use "condition"
predict_response(fit, terms = "c12hour [40:60]", condition = c(neg_c_7 = 20))

# to plot ggeffects-objects, you can use the 'plot()'-function.
# the following examples show how to build your ggplot by hand.
```

```
# plot predicted values, remaining covariates held constant
library(ggplot2)
mydf <- predict_response(fit, terms = "c12hour")
ggplot(mydf, aes(x, predicted)) +
  geom_line() +
  geom_ribbon(aes(ymin = conf.low, ymax = conf.high), alpha = 0.1)

# three variables, so we can use facets and groups
mydf <- predict_response(fit, terms = c("c12hour", "c161sex", "c172code"))
ggplot(mydf, aes(x = x, y = predicted, colour = group)) +
  stat_smooth(method = "lm", se = FALSE) +
  facet_wrap(~facet, ncol = 2)

# select specific levels for grouping terms
mydf <- predict_response(fit, terms = c("c12hour", "c172code [1,3]", "c161sex"))
ggplot(mydf, aes(x = x, y = predicted, colour = group)) +
  stat_smooth(method = "lm", se = FALSE) +
  facet_wrap(~facet) +
  labs(
    y = get_y_title(mydf),
    x = get_x_title(mydf),
    colour = get_legend_title(mydf)
  )

# level indication also works for factors with non-numeric levels
# and in combination with numeric levels for other variables
data(efc)
efc$c172code <- sjlabelled::as_label(efc$c172code)
fit <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)
predict_response(fit, terms = c("c12hour",
  "c172code [low level of education, high level of education]",
  "c161sex [1]"))

# when "terms" is a named list
predict_response(fit, terms = list(
  c12hour = seq(0, 170, 30),
  c172code = c("low level of education", "high level of education"),
  c161sex = 1)
)

# use categorical value on x-axis, use axis-labels, add error bars
dat <- predict_response(fit, terms = c("c172code", "c161sex"))
ggplot(dat, aes(x, predicted, colour = group)) +
  geom_point(position = position_dodge(0.1)) +
  geom_errorbar(
    aes(ymin = conf.low, ymax = conf.high),
    position = position_dodge(0.1)
  ) +
  scale_x_discrete(breaks = 1:3, labels = get_x_labels(dat))

# 3-way-interaction with 2 continuous variables
data(efc)
# make categorical
```

```
efc$c161sex <- as_factor(efc$c161sex)
fit <- lm(neg_c_7 ~ c12hour * barthtot * c161sex, data = efc)
# select only levels 30, 50 and 70 from continuous variable Barthel-Index
dat <- predict_response(fit, terms = c("c12hour", "barthtot [30,50,70]", "c161sex"))
ggplot(dat, aes(x = x, y = predicted, colour = group)) +
  stat_smooth(method = "lm", se = FALSE, fullrange = TRUE) +
  facet_wrap(~facet) +
  labs(
    colour = get_legend_title(dat),
    x = get_x_title(dat),
    y = get_y_title(dat),
    title = get_title(dat)
  )

# or with ggeffects' plot-method
plot(dat, ci = FALSE)


# predictions for polynomial terms
data(efc)
fit <- glm(
  tot_sc_e ~ c12hour + e42dep + e17age + I(e17age^2) + I(e17age^3),
  data = efc,
  family = poisson()
)
predict_response(fit, terms = "e17age")
```

`pretty_range` *Create a pretty sequence over a range of a vector*

Description

Creates an evenly spaced, pretty sequence of numbers for a range of a vector.

Usage

```
pretty_range(x, n = NULL, length = NULL)
```

Arguments

`x`	A numeric vector.
`n`	Numeric value, indicating the size of how many values are used to create a pretty sequence. If `x` has a large value range (> 100), `n` could be something between 1 to 5. If `x` has a rather small amount of unique values, `n` could be something between 10 to 20. If `n = NULL`, `pretty_range()` automatically tries to find a pretty sequence.
`length`	Integer value, as alternative to `n`, defines the number of intervals to be returned.

Value

A numeric vector with a range corresponding to the minimum and maximum values of `x`. If `x` is missing, a function, pre-programmed with `n` and `length` is returned. See examples.

Examples

```
data(iris)
# pretty range for vectors with decimal points
pretty_range(iris$Petal.Length)

# pretty range for large range, increasing by 50
pretty_range(1:1000)

# increasing by 20
pretty_range(1:1000, n = 7)

# return 10 intervals
pretty_range(1:1000, length = 10)

# same result
pretty_range(1:1000, n = 2.5)

# function factory
range_n_5 <- pretty_range(n = 5)
range_n_5(1:1000)
```

`residualize_over_grid` *Compute partial residuals from a data grid*

Description

This function computes partial residuals based on a data grid, where the data grid is usually a data frame from all combinations of factor variables or certain values of numeric vectors. This data grid is usually used as `newdata` argument in `predict()`, and can be created with `new_data()`.

Usage

```
residualize_over_grid(grid, model, ...)

## S3 method for class 'data.frame'
residualize_over_grid(grid, model, predictor_name, ...)

## S3 method for class 'ggeffects'
residualize_over_grid(grid, model, protect_names = TRUE, ...)
```

Arguments

grid	A data frame representing the data grid, or an object of class ggeffects, as returned by predict_response().
model	The model for which to compute partial residuals. The data grid grid should match to predictors in the model.
...	Currently not used.
predictor_name	The name of the focal predictor, for which partial residuals are computed.
protect_names	Logical, if TRUE, preserves column names from the ggeffects objects that is used as grid.

Value

A data frame with residuals for the focal predictor.

Partial Residuals

For **generalized linear models** (glms), residualized scores are computed as inv.link(link(Y) + r) where Y are the predicted values on the response scale, and r are the *working* residuals.

For (generalized) linear **mixed models**, the random effect are also partialled out.

References

Fox J, Weisberg S. Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. Journal of Statistical Software 2018;87.

Examples

```
library(ggeffects)
set.seed(1234)
x <- rnorm(200)
z <- rnorm(200)
# quadratic relationship
y <- 2 * x + x^2 + 4 * z + rnorm(200)

d <- data.frame(x, y, z)
model <- lm(y ~ x + z, data = d)

pr <- predict_response(model, c("x [all]", "z"))
head(residualize_over_grid(pr, model))
```

test_predictions *(Pairwise) comparisons between predictions (marginal effects)*

Description

Function to test differences of adjusted predictions for statistical significance. This is usually called contrasts or (pairwise) comparisons, or "marginal effects". `hypothesis_test()` is an alias.

Usage

```
test_predictions(model, ...)

hypothesis_test(model, ...)

## Default S3 method:
test_predictions(
  model,
  terms = NULL,
  by = NULL,
  test = "pairwise",
  equivalence = NULL,
  scale = "response",
  p_adjust = NULL,
  df = NULL,
  ci_level = 0.95,
  collapse_levels = FALSE,
  verbose = TRUE,
  ci.lvl = ci_level,
  ...
)

## S3 method for class 'ggeffects'
test_predictions(
  model,
  by = NULL,
  test = "pairwise",
  equivalence = NULL,
  scale = "response",
  p_adjust = NULL,
  df = NULL,
  collapse_levels = FALSE,
  verbose = TRUE,
  ...
)
```

Arguments

`model`	A fitted model object, or an object of class `ggeffects`.

`...` Arguments passed down to `data_grid()` when creating the reference grid and to `marginaleffects::predictions()` resp. `marginaleffects::slopes()`. For instance, arguments `type` or `transform` can be used to back-transform comparisons and contrasts to different scales. `vcov` can be used to calculate heteroscedasticity-consistent standard errors for contrasts. See examples at the bottom of this vignette for further details. To define a heteroscedasticity-consistent variance-covariance matrix, you can either use the same arguments as for `predict_response()` etc., namely `vcov_fun`, `vcov_type` and `vcov_args`. These are then transformed into a matrix and passed down to the `vcov` argument in *marginaleffects*. Or you directly use the `vcov` argument. See `?marginaleffects::slopes` for further details.

`terms` Character vector with the names of the focal terms from `model`, for which contrasts or comparisons should be displayed. At least one term is required, maximum length is three terms. If the first focal term is numeric, contrasts or comparisons for the *slopes* of this numeric predictor are computed (possibly grouped by the levels of further categorical focal predictors).

`by` Character vector specifying the names of predictors to condition on. Hypothesis test is then carried out for focal terms by each level of `by` variables. This is useful especially for interaction terms, where we want to test the interaction within "groups". `by` is only relevant for categorical predictors.

`test` Hypothesis to test. By default, pairwise-comparisons are conducted. See section *Introduction into contrasts and pairwise comparisons*.

`equivalence` ROPE's lower and higher bounds. Should be `"default"` or a vector of length two (e.g., `c(-0.1, 0.1)`). If `"default"`, `bayestestR::rope_range()` is used. Instead of using the `equivalence` argument, it is also possible to call the `equivalence_test()` method directly. This requires the **parameters** package to be loaded. When using `equivalence_test()`, two more columns with information about the ROPE coverage and decision on H0 are added. Furthermore, it is possible to `plot()` the results from `equivalence_test()`. See `bayestestR::equivalence_test()` resp. `parameters::equivalence_test.lm()` for details.

`scale` Character string, indicating the scale on which the contrasts or comparisons are represented. Can be one of:

- `"response"` (default), which would return contrasts on the response scale (e.g. for logistic regression, as probabilities);
- `"link"` to return contrasts on scale of the linear predictors (e.g. for logistic regression, as log-odds);
- `"probability"` (or `"probs"`) returns contrasts on the probability scale, which is required for some model classes, like `MASS::polr()`;
- `"oddsratios"` to return contrasts on the odds ratio scale (only applies to logistic regression models);
- `"irr"` to return contrasts on the odds ratio scale (only applies to count models);
- or a transformation function like `"exp"` or `"log"`, to return transformed (exponentiated respectively logarithmic) contrasts; note that these transformations are applied to the *response scale*.

Note: If the scale argument is not supported by the provided model, it is automaticaly changed to a supported scale-type (a message is printed when verbose = TRUE).

p_adjust — Character vector, if not NULL, indicates the method to adjust p-values. See stats::p.adjust() or stats::p.adjust.methods for details. Further possible adjustment methods are "tukey" or "sidak", and for johnson_neyman(), "fdr" (or "bh") and "esarey" (or its short-cut "es") are available options. Some caution is necessary when adjusting p-value for multiple comparisons. See also section *P-value adjustment* below.

df — Degrees of freedom that will be used to compute the p-values and confidence intervals. If NULL, degrees of freedom will be extracted from the model using insight::get_df() with type = "wald".

ci_level — Numeric, the level of the confidence intervals.

collapse_levels — Logical, if TRUE, term labels that refer to identical levels are no longer separated by "-", but instead collapsed into a unique term label (e.g., "level a-level a" becomes "level a"). See 'Examples'.

verbose — Toggle messages and warnings.

ci.lvl — Deprecated, please use ci_level.

Value

A data frame containing predictions (e.g. for test = NULL), contrasts or pairwise comparisons of adjusted predictions or estimated marginal means.

Introduction into contrasts and pairwise comparisons

There are many ways to test contrasts or pairwise comparisons. A detailed introduction with many (visual) examples is shown in this vignette.

P-value adjustment for multiple comparisons

Note that p-value adjustment for methods supported by p.adjust() (see also p.adjust.methods), each row is considered as one set of comparisons, no matter which test was specified. That is, for instance, when test_predictions() returns eight rows of predictions (when test = NULL), and p_adjust = "bonferroni", the p-values are adjusted in the same way as if we had a test of pairwise comparisons (test = "pairwise") where eight rows of comparisons are returned. For methods "tukey" or "sidak", a rank adjustment is done based on the number of combinations of levels from the focal predictors in terms. Thus, the latter two methods may be useful for certain tests only, in particular pairwise comparisons.

For johnson_neyman(), the only available adjustment methods are "fdr" (or "bh") (*Benjamini & Hochberg (1995)*) and "esarey" (or "es") (*Esarey and Sumner 2017*). These usually return similar results. The major difference is that "fdr" can be slightly faster and more stable in edge cases, however, confidence intervals are not updated. Only the p-values are adjusted. "esarey" is slower, but confidence intervals are updated as well.

Global Options to Customize Tables when Printing

The `verbose` argument can be used to display or silence messages and warnings. Furthermore, `options()` can be used to set defaults for the `print()` and `print_html()` method. The following options are available, which can simply be run in the console:

- `ggeffects_ci_brackets`: Define a character vector of length two, indicating the opening and closing parentheses that encompass the confidence intervals values, e.g. `options(ggeffects_ci_brackets = c("[", "]"))`.
- `ggeffects_collapse_ci`: Logical, if `TRUE`, the columns with predicted values (or contrasts) and confidence intervals are collapsed into one column, e.g. `options(ggeffects_collapse_ci = TRUE)`.
- `ggeffects_collapse_p`: Logical, if `TRUE`, the columns with predicted values (or contrasts) and p-values are collapsed into one column, e.g. `options(ggeffects_collapse_p = TRUE)`. Note that p-values are replaced by asterisk-symbols (stars) or empty strings when `ggeffects_collapse_p = TRUE`, depending on the significance level.
- `ggeffects_collapse_tables`: Logical, if `TRUE`, multiple tables for subgroups are combined into one table. Only works when there is more than one focal term, e.g. `options(ggeffects_collapse_tables = TRUE)`.
- `ggeffects_output_format`: String, either `"text"`, `"markdown"` or `"html"`. Defines the default output format from `predict_response()`. If `"html"`, a formatted HTML table is created and printed to the view pane. `"markdown"` creates a markdown-formatted table inside Rmarkdown documents, and prints a text-format table to the console when used interactively. If `"text"` or `NULL`, a formatted table is printed to the console, e.g. `options(ggeffects_output_format = "html")`.
- `ggeffects_html_engine`: String, either `"tt"` or `"gt"`. Defines the default engine to use for printing HTML tables. If `"tt"`, the *tinytable* package is used, if `"gt"`, the *gt* package is used, e.g. `options(ggeffects_html_engine = "gt")`.

Use `options(<option_name> = NULL)` to remove the option.

References

Esarey, J., & Sumner, J. L. (2017). Marginal effects in interaction models: Determining and controlling the false positive rate. Comparative Political Studies, 1–33. Advance online publication. doi: 10.1177/0010414017730080

See Also

There is also an `equivalence_test()` method in the **parameters** package (`parameters::equivalence_test.lm()`), which can be used to test contrasts or comparisons for practical equivalence. This method also has a `plot()` method, hence it is possible to do something like:

```
library(parameters)
predict_response(model, focal_terms) |>
  equivalence_test() |>
  plot()
```

Examples

```
data(efc)
efc$c172code <- as.factor(efc$c172code)
efc$c161sex <- as.factor(efc$c161sex)
levels(efc$c161sex) <- c("male", "female")
m <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)

# direct computation of comparisons
test_predictions(m, "c172code")

# passing a `ggeffects` object
pred <- predict_response(m, "c172code")
test_predictions(pred)

# test for slope
test_predictions(m, "c12hour")

# interaction - contrasts by groups
m <- lm(barthtot ~ c12hour + c161sex * c172code + neg_c_7, data = efc)
test_predictions(m, c("c161sex", "c172code"), test = NULL)

# interaction - pairwise comparisons by groups
test_predictions(m, c("c161sex", "c172code"))

# equivalence testing
test_predictions(m, c("c161sex", "c172code"), equivalence = c(-2.96, 2.96))

# equivalence testing, using the parameters package
pr <- predict_response(m, c("c161sex", "c172code"))
parameters::equivalence_test(pr)

# interaction - collapse unique levels
test_predictions(m, c("c161sex", "c172code"), collapse_levels = TRUE)

# p-value adjustment
test_predictions(m, c("c161sex", "c172code"), p_adjust = "tukey")

# not all comparisons, only by specific group levels
test_predictions(m, "c172code", by = "c161sex")

# specific comparisons
test_predictions(m, c("c161sex", "c172code"), test = "b2 = b1")

# interaction - slope by groups
m <- lm(barthtot ~ c12hour + neg_c_7 * c172code + c161sex, data = efc)
test_predictions(m, c("neg_c_7", "c172code"))

# Example: marginal effects ----------------------------
# ------------------------------------------------------
data(iris)
```

```
m <- lm(Petal.Width ~ Petal.Length + Species, data = iris)

# we now want the marginal effects for "Species". We can calculate
# the marginal effect using the "margins" package
margins::margins(m, variables = "Species")

# we get the same marginal effect from the "marginaleffects" package
marginaleffects::avg_slopes(m, variables = "Species")

# finally, test_predictions() returns the same. while the previous results
# report the marginal effect compared to the reference level "setosa",
# test_predictions() returns the marginal effects for all pairwise comparisons
test_predictions(m, "Species")
```

`values_at` *Calculate representative values of a vector*

Description

This function calculates representative values of a vector, like minimum/maximum values or lower, median and upper quartile etc., which can be used for numeric vectors to plot adjusted predictions at these representative values.

Usage

```
values_at(x, values = "meansd")

representative_values(x, values = "meansd")
```

Arguments

`x`	A numeric vector.
`values`	Character vector, naming a pattern for which representative values should be calculcated. • `"minmax"`: (default) minimum and maximum values (lower and upper bounds) of the moderator are used to plot the interaction between independent variable and moderator. • `"meansd"`: uses the mean value of the moderator as well as one standard deviation below and above mean value to plot the effect of the moderator on the independent variable. • `"zeromax"`: is similar to the `"minmax"` option, however, `0` is always used as minimum value for the moderator. This may be useful for predictors that don't have an empirical zero-value, but absence of moderation should be simulated by using 0 as minimum. • `"fivenum"`: calculates and uses the Tukey's five number summary (minimum, lower-hinge, median, upper-hinge, maximum) of the moderator value.

- "quart": calculates and uses the quartiles (lower, median and upper) of the moderator value, *including* minimum and maximum value.
- "quart2": calculates and uses the quartiles (lower, median and upper) of the moderator value, *excluding* minimum and maximum value.
- "terciles": calculates and uses the terciles (lower and upper third) of the moderator value, *including* minimum and maximum value.
- "terciles2": calculates and uses the terciles (lower and upper third) of the moderator value, *excluding* minimum and maximum value.
- "all": uses all values of the moderator variable. Note that this option only applies to type = "eff", for numeric moderator values.

Value

A numeric vector of length two or three, representing the required values from x, like minimum/maximum value or mean and +/- 1 SD. If x is missing, a function, pre-programmed with n and length is returned. See examples.

Examples

```
data(efc)
values_at(efc$c12hour)
values_at(efc$c12hour, "quart2")

mean_sd <- values_at(values = "meansd")
mean_sd(efc$c12hour)
```

---

vcov *Calculate variance-covariance matrix for adjusted predictions*

Description

Returns the variance-covariance matrix for the predicted values from object.

Usage

```
## S3 method for class 'ggeffects'
vcov(
  object,
  vcov_fun = NULL,
  vcov_type = NULL,
  vcov_args = NULL,
  vcov.fun = vcov_fun,
  vcov.type = vcov_type,
  vcov.args = vcov_args,
  ...
)
```

Arguments

object	An object of class "ggeffects", as returned by predict_response().
vcov_fun	Variance-covariance matrix used to compute uncertainty estimates (e.g., for confidence intervals based on robust standard errors). This argument accepts a covariance matrix, a function which returns a covariance matrix, or a string which identifies the function to be used to compute the covariance matrix. • A (variance-covariance) matrix • A function which returns a covariance matrix (e.g., stats::vcov()) • A string which indicates the estimation type for the heteroscedasticity-consistent variance-covariance matrix, e.g. vcov_fun = "HC0". Possible values are "HC0", "HC1", "HC2", "HC3", "HC4", "HC4m", and "HC5", which will then call the vcovHC()-function from the **sandwich** package, using the specified type. Further possible values are "CR0", "CR1", "CR1p", "CR1S", "CR2", and "CR3", which will call the vcovCR()-function from the **clubSandwich** package. • A string which indicates the name of the vcov*()-function from the **sandwich** or **clubSandwich** packages, e.g. vcov_fun = "vcovCL", which is used to compute (cluster) robust standard errors for predictions. If NULL, standard errors (and confidence intervals) for predictions are based on the standard errors as returned by the predict()-function. **Note** that probably not all model objects that work with ggpredict() are also supported by the **sandwich** or **clubSandwich** packages. See details in this vignette.
vcov_type	Character vector, specifying the estimation type for the robust covariance matrix estimation (see ?sandwich::vcovHC or ?clubSandwich::vcovCR for details). Only used when vcov_fun is a character string indicating one of the functions from those packages.
vcov_args	List of named vectors, used as additional arguments that are passed down to vcov_fun.
vcov.fun, vcov.type, vcov.args	Deprecated. Use vcov_fun, vcov_type and vcov_args instead.
...	Currently not used.

Details

The returned matrix has as many rows (and columns) as possible combinations of predicted values from the predict_response() call. For example, if there are two variables in the terms-argument of predict_response() with 3 and 4 levels each, there will be 3*4 combinations of predicted values, so the returned matrix has a 12x12 dimension. In short, nrow(object) is always equal to nrow(vcov(object)). See also 'Examples'.

Value

The variance-covariance matrix for the predicted values from object.

Examples

```
data(efc)
model <- lm(barthtot ~ c12hour + neg_c_7 + c161sex + c172code, data = efc)
result <- predict_response(model, c("c12hour [meansd]", "c161sex"))

vcov(result)

# compare standard errors
sqrt(diag(vcov(result)))
as.data.frame(result)

# only two predicted values, no further terms
# vcov() returns a 2x2 matrix
result <- predict_response(model, "c161sex")
vcov(result)

# 2 levels for c161sex multiplied by 3 levels for c172code
# result in 6 combinations of predicted values
# thus vcov() returns a 6x6 matrix
result <- predict_response(model, c("c161sex", "c172code"))
vcov(result)
```

Index