Package 'vpc'

July 31, 2017

Title Create visual predictive checks in R

Version 0.9.1

Date 2017-07-25

Author Ron Keizer <ronkeizer@gmail.com>

Maintainer Ron Keizer <ronkeizer@gmail.com>

Description Visual predictive checks are a commonly used diagnostic plot in pharmacometrics, showing how percentiles for observated data compare to those same percentiles of data simulated from a model. This R package allows creation of VPCs without the use of PsN and Xpose4.

Depends R (>= 3.1.0)

Imports classInt, dplyr, reshape2, MASS, survival, ggplot2, readr

License MIT + file LICENSE

LazyData true

URL https://github.com/ronkeizer/vpc

Suggests knitr, testit

VignetteBuilder knitr

RoxygenNote 6.0.1

R topics documented:

`vpc-package` *VPC package*

Description

Create Visual Predictive Checks in R

Author(s)

Ron Keizer `<ronkeizer@gmail.com>`

`add_noise` *Add noise / residual error to data*

Description

Add noise / residual error to data

Usage

```
add_noise(x, ruv = list(proportional = 0, additive = 0, exponential = 0))
```

Arguments

`x`	data
`ruv`	list describing the magnitude of errors. List arguments: "proportional", "additive", "exponential".

add_sim_index_number *Add sim index number*

Description

Add simulation index number to simulation when not present

Usage

```
add_sim_index_number(sim, id = "id")
```

Arguments

sim	a data.frame containing the simulation data
id	character specifying the column name in the data.frame

auto_bin *Calculate appropriate bin separators for vpc*

Description

Calculate appropriate bin separators for vpc

Usage

```
auto_bin(dat, type = "kmeans", n_bins = 8, verbose = FALSE)
```

Arguments

dat	data frame
type	auto-binning type: "density", "time", or "data"
n_bins	number of bins to use. For "density" the function might not return a solution with the exact number of bins.
verbose	show warnings and other messages (TRUE or FALSE)

Details

This function calculates bin separators (e.g. for use in a vpc) based on nadirs in the density functions for the indenpendent variable

Value

A vector of bin separators

See Also

vpc

`bin_data` *Function to bin data based on a vector of bin separators, e.g. for use in VPC*

Description

Function to bin data based on a vector of bin separators, e.g. for use in VPC

Usage

```
bin_data(x, bins = c(0, 3, 5, 7), idv = "time")
```

Arguments

x	data
bins	numeric vector specifying bin separators
idv	variable in the data specifies the independent variable (e.g. "time")

`create_vpc_theme` *Create new vpc theme*

Description

Create new vpc theme

Usage

```
create_vpc_theme(...)
```

Arguments

...	pass arguments to ‘new_vpc_theme‘

draw_params_mvr *Draw parameters from multivariate distribution*

Description

Draw parameters from multivariate distribution

Usage

```
draw_params_mvr(ids, n_sim, theta, omega_mat, par_names = NULL)
```

Arguments

ids	vector of ids
n_sim	number of simulations
theta	theta vector
omega_mat	omega matrix
par_names	parameter names vector

new_vpc_theme *Create a customized VPC theme*

Description

Create a customized VPC theme

Usage

```
new_vpc_theme(update = NULL)
```

Arguments

update	list containing the plot elements to be updated. Run 'new_vpc_theme()' with no arguments to show an overview of available plot elements.

Details

This function creates a theme that customizes how the VPC looks, i.e. colors, fills, transparencies, linetypes an sizes, etc. The following arguments can be specified in the input list:

- obs_color: color for observationss points
- obs_size: size for observation points
- obs_median_color: color for median observation line
- obs_median_linetype: linetype for median observation line

- obs_median_size: size for median observation line
- obs_ci_color: color for observation CI lines
- obs_ci_linetype: linetype for observation CI lines
- obs_ci_size: size for observations CI lines
- sim_pi_fill: fill color for simulated prediction interval areas
- sim_pi_alpha: transparency for simulated prediction interval areas
- sim_pi_color: color for simulated prediction interval lines
- sim_pi_linetype: linetype for simulated prediction interval lines
- sim_pi_size: size for simulated prediction interval lines
- sim_median_fill: fill color for simulated median area
- sim_median_alpha: transparency for simulated median area
- sim_median_color: color for simulated median line
- sim_median_linetype: linetype for simulated median line
- sim_median_size: size for simulated median line
- bin_separators_color: color for bin separator lines, NA for don't plot
- bin_separators_location: where to plot bin separators ("t" for top, "b" for bottom)

Value

A list with vpc theme specifiers

`pk_oral_1cmt` *Simulate PK data from a 1-compartment oral model*

Description

Simulate PK data from a 1-compartment oral model

Usage

```
pk_oral_1cmt(t, tau = 24, dose = 120, ka = 1, ke = 1, cl = 10,
  ruv = NULL)
```

Arguments

`t`	Time after dose
`tau`	Dosing interval
`dose`	Dose
`ka`	Absorption rate
`ke`	Elimination rate
`cl`	Clearance
`ruv`	Residual variability

Value

A vector of predicted values, with or without added residual variability

---

`plot_vpc` *VPC plotting function*

Description

VPC plotting function

Usage

```
plot_vpc(db, show = NULL, vpc_theme = NULL, smooth = TRUE,
  log_x = FALSE, log_y = FALSE, title = NULL, xlab = "Time",
  ylab = "Dependent value", verbose = FALSE)
```

Arguments

`db`	object created using the ‘vpc‘ function
`show`	what to show in VPC (obs_dv, obs_ci, pi, pi_as_area, pi_ci, obs_median, sim_median, sim_median_ci)
`vpc_theme`	theme to be used in VPC. Expects list of class vpc_theme created with function vpc_theme()
`smooth`	"smooth" the VPC (connect bin midpoints) or show bins as rectangular boxes. Default is TRUE.
`log_x`	Boolean indicting whether x-axis should be shown as logarithmic. Default is FALSE.
`log_y`	Boolean indicting whether y-axis should be shown as logarithmic. Default is FALSE.
`title`	title
`xlab`	ylab as numeric vector of size 2
`ylab`	ylab as numeric vector of size 2
`verbose`	verbosity (T/F)

`read_table_nm` *NONMEM output table import function*

Description

Quickly import NONMEM output tables into R. Function taken from 'modelviz' package by Benjamin Guiastrennec. When both `skip` and `header` are `NULL`, `read_nmtab` will automatically detect the optimal settings to import the tables. When more than one files are provided for a same NONMEM run, they will be combined into a single `data.frame`.

Usage

```
read_table_nm(file = NULL, skip = NULL, header = NULL,
  rm_duplicates = FALSE, nonmem_tab = TRUE)
```

Arguments

`file`	full file name
`skip`	number of lines to skip before reading data
`header`	logical value indicating whether the file contains the names of the variables as its first line
`rm_duplicates`	logical value indicating whether duplicated columns should be removed
`nonmem_tab`	logical value indicading to the function whether the file is a table or a nonmem additional output file.

Value

A `data.frame`

Examples

```
## Not run:
data <- read_table_nm(file = '../models/pk/sdtab101')

## End(Not run)
```

replace_list_elements *Replace list elements by name*

Description

Replace list elements by name

Usage

```
replace_list_elements(list, replacement)
```

Arguments

list	original list
replacement	replacement list

Details

Finds and replaces list elements by name and throws an error if an element is not available in the original list. This is a local duplicate of the PKPDmisc copy for the VPC package to reduce dependency on PKPDmisc at this time.

Examples

```
## Not run:
list <- list(ipred = "ipred", dv = "dv", idv = "idv", "pred" = "pred")
replacement <- list(dv = "conc", idv = "time")
list <- replace_list_elements(list, replacement)

## End(Not run)
```

rtte_obs_nm *Simulated RTTE data (1x)*

Description

An example dataset with simulated repeated time-to-event data

Usage

```
rtte_obs_nm
```

Format

An object of class `data.frame` with 573 rows and 6 columns.

rtte_sim_nm *Simulated RTTE data (100x)*

Description

An example dataset with simulated repeated time-to-event data (100 simulations)

Usage

```
rtte_sim_nm
```

Format

An object of class `data.frame` with 2000000 rows and 7 columns.

show_default *Defaults for show argument*

Description

Defaults for show argument

Usage

```
show_default
```

Format

An object of class `list` of length 11.

simple_data *A small rich dataset*

Description

A small rich dataset

Usage

```
simple_data
```

Format

An object of class `list` of length 2.

Details

a list containing the obs and sim data for an example dataset to run a simple vpc.

Examples

```
## Not run:
vpc(simple_data$sim, simple_data$obs)

## End(Not run)
```

`sim_data` *Simulate data based on a model and parameter distributions*

Description

Simulate data based on a model and parameter distributions

Usage

```
sim_data(design = cbind(id = c(1, 1, 1), idv = c(0, 1, 2)),
  model = function(x) {     return(x$alpha + x$beta) }, theta, omega_mat,
  par_names, par_values = NULL, draw_iiv = "mvrnorm",
  error = list(proportional = 0, additive = 0, exponential = 0), n = 100)
```

Arguments

`design`	a design dataset. See example
`model`	A function with the first argument the simulation design, i.e. a dataset with the columns ... The second argument to this function is a dataset with parameters for every individual. This can be supplied by the user, or generated by this sim_data if theta and omega_mat are supplied.
`theta`	vector of fixed effect parameters
`omega_mat`	vector of between subject random effects, specified as lower triangle
`par_names`	A character vector linking the parameters in the model to the variables in the dataset. See example.
`par_values`	parameter values
`draw_iiv`	draw between subject random effects?
`error`	see example
`n`	number of simulations to perform

Details

This function generates the simulated dependent values for use in the VPC plotting function.

Value

a vector of simulated dependent variables (for us in the VPC plotting function)

See Also

vpc

---

theme_empty *Empty theme*

Description

Empty theme

Usage

```
theme_empty()
```

---

theme_plain *Nicer default theme for ggplot2*

Description

Nicer default theme for ggplot2

Usage

```
theme_plain()
```

---

triangle_to_full *Lower to full triangle*

Description

Convert the lower triangle of a covariance matrix to a full matrix object

Usage

```
triangle_to_full(vect)
```

Arguments

vect	the lower triangle of a covariance matrix

vpc *VPC function*

Description

Creates a VPC plot from observed and simulation data

Usage

```
vpc(sim = NULL, obs = NULL, psn_folder = NULL, bins = "jenks",
  n_bins = "auto", bin_mid = "mean", obs_cols = NULL, sim_cols = NULL,
  software = "auto", show = NULL, stratify = NULL, pred_corr = FALSE,
  pred_corr_lower_bnd = 0, pi = c(0.05, 0.95), ci = c(0.05, 0.95),
  uloq = NULL, lloq = NULL, log_y = FALSE, log_y_min = 0.001,
  xlab = NULL, ylab = NULL, title = NULL, facet_names = TRUE,
  smooth = TRUE, vpc_theme = NULL, facet = "wrap", labeller = NULL,
  vpcdb = FALSE, verbose = FALSE)
```

Arguments

sim	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
obs	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
psn_folder	instead of specyfing "sim" and "obs", specify a PsN-generated VPC-folder
bins	either "density", "time", or "data", "none", or one of the approaches available in classInterval() such as "jenks" (default) or "pretty", or a numeric vector specifying the bin separators.
n_bins	when using the "auto" binning method, what number of bins to aim for
bin_mid	either "mean" for the mean of all timepoints (default) or "middle" to use the average of the bin boundaries.
obs_cols	observation dataset column names (list elements: "dv", "idv", "id", "pred")
sim_cols	simulation dataset column names (list elements: "dv", "idv", "id", "pred")
software	name of software platform using (eg nonmem, phoenix)
show	what to show in VPC (obs_dv, obs_ci, pi, pi_as_area, pi_ci, obs_median, sim_median, sim_median_ci)
stratify	character vector of stratification variables. Only 1 or 2 stratification variables can be supplied.
pred_corr	perform prediction-correction?
pred_corr_lower_bnd	lower bound for the prediction-correction

pi	simulated prediction interval to plot. Default is c(0.05, 0.95),
ci	confidence interval to plot. Default is (0.05, 0.95)
uloq	Number or NULL indicating upper limit of quantification. Default is NULL.
lloq	Number or NULL indicating lower limit of quantification. Default is NULL.
log_y	Boolean indicting whether y-axis should be shown as logarithmic. Default is FALSE.
log_y_min	minimal value when using log_y argument. Default is 1e-3.
xlab	ylab as numeric vector of size 2
ylab	ylab as numeric vector of size 2
title	title
smooth	"smooth" the VPC (connect bin midpoints) or show bins as rectangular boxes. Default is TRUE.
vpc_theme	theme to be used in VPC. Expects list of class vpc_theme created with function vpc_theme()
facet	either "wrap", "columns", or "rows"
labeller	ggplot2 labeller function to be passed to underlying ggplot object
vpcdb	Boolean whether to return the underlying vpcdb rather than the plot
verbose	show debugging information (TRUE or FALSE)

Value

a list containing calculated VPC information (when vpcdb=TRUE), or a ggplot2 object (default)

See Also

sim_data, vpc_cens, vpc_tte

---

vpc_cat *VPC function for categorical*

Description

Creates a VPC plot from observed and simulation data sim,

Usage

```
vpc_cat(sim = NULL, obs = NULL, psn_folder = NULL, bins = "jenks",
  n_bins = "auto", bin_mid = "mean", obs_cols = NULL, sim_cols = NULL,
  software = "auto", show = NULL, ci = c(0.05, 0.95), uloq = NULL,
  lloq = NULL, xlab = NULL, ylab = NULL, title = NULL, smooth = TRUE,
  stratify = NULL, stratify_color = NULL, vpc_theme = NULL,
  facet = "wrap", labeller = NULL, plot = TRUE, vpcdb = FALSE,
  verbose = FALSE)
```

Arguments

sim	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
obs	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
psn_folder	instead of specyfing "sim" and "obs", specify a PsN-generated VPC-folder
bins	either "density", "time", or "data", "none", or one of the approaches available in classInterval() such as "jenks" (default) or "pretty", or a numeric vector specifying the bin separators.
n_bins	when using the "auto" binning method, what number of bins to aim for
bin_mid	either "mean" for the mean of all timepoints (default) or "middle" to use the average of the bin boundaries.
obs_cols	observation dataset column names (list elements: "dv", "idv", "id", "pred")
sim_cols	simulation dataset column names (list elements: "dv", "idv", "id", "pred")
software	name of software platform using (eg nonmem, phoenix)
show	what to show in VPC (obs_ci, pi, pi_as_area, pi_ci, obs_median, sim_median, sim_median_ci)
ci	confidence interval to plot. Default is (0.05, 0.95)
uloq	Number or NULL indicating upper limit of quantification. Default is NULL.
lloq	Number or NULL indicating lower limit of quantification. Default is NULL.
xlab	ylab as numeric vector of size 2
ylab	ylab as numeric vector of size 2
title	title
smooth	"smooth" the VPC (connect bin midpoints) or show bins as rectangular boxes. Default is TRUE.
stratify	character vector of stratification variables. Only 1 or 2 stratification variables can be supplied.
stratify_color	variable to stratify and color lines for observed data. Only 1 stratification variables can be supplied.
vpc_theme	theme to be used in VPC. Expects list of class vpc_theme created with function vpc_theme()
facet	either "wrap", "columns", or "rows"
labeller	ggplot2 labeller function to be passed to underlying ggplot object
plot	Boolean indicting whether to plot the ggplot2 object after creation. Default is FALSE.
vpcdb	boolean whether to return the underlying vpcdb rather than the plot
verbose	show debugging information (TRUE or FALSE)

Value

a list containing calculated VPC information (when vpcdb=TRUE), or a ggplot2 object (default)

See Also

vpc

`vpc_cens` *VPC function for left- or right-censored data (e.g. BLOQ data)*

Description

Creates a VPC plot from observed and simulation data sim,

Usage

```
vpc_cens(sim = NULL, obs = NULL, psn_folder = NULL, bins = "jenks",
  n_bins = 8, bin_mid = "mean", obs_cols = NULL, sim_cols = NULL,
  software = "auto", show = NULL, stratify = NULL,
  stratify_color = NULL, ci = c(0.05, 0.95), uloq = NULL, lloq = NULL,
  plot = FALSE, xlab = NULL, ylab = NULL, title = NULL, smooth = TRUE,
  vpc_theme = NULL, facet = "wrap", labeller = NULL, vpcdb = FALSE,
  verbose = FALSE)
```

Arguments

`sim`	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
`obs`	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
`psn_folder`	instead of specyfing "sim" and "obs", specify a PsN-generated VPC-folder
`bins`	either "density", "time", or "data", or a numeric vector specifying the bin separators.
`n_bins`	number of bins
`bin_mid`	either "mean" for the mean of all timepoints (default) or "middle" to use the average of the bin boundaries.
`obs_cols`	observation dataset column names (list elements: "dv", "idv", "id", "pred")
`sim_cols`	simulation dataset column names (list elements: "dv", "idv", "id", "pred")
`software`	name of software platform using (eg nonmem, phoenix)
`show`	what to show in VPC (obs_ci, pi, pi_as_area, pi_ci, obs_median, sim_median, sim_median_ci)

stratify	character vector of stratification variables. Only 1 or 2 stratification variables can be supplied.
stratify_color	variable to stratify and color lines for observed data. Only 1 stratification variables can be supplied.
ci	confidence interval to plot. Default is (0.05, 0.95)
uloq	Number or NULL indicating upper limit of quantification. Default is NULL.
lloq	Number or NULL indicating lower limit of quantification. Default is NULL.
plot	Boolean indacting whether to plot the ggplot2 object after creation. Default is FALSE.
xlab	ylab as numeric vector of size 2
ylab	ylab as numeric vector of size 2
title	title
smooth	"smooth" the VPC (connect bin midpoints) or show bins as rectangular boxes. Default is TRUE.
vpc_theme	theme to be used in VPC. Expects list of class vpc_theme created with function vpc_theme()
facet	either "wrap", "columns", or "rows"
labeller	ggplot2 labeller function to be passed to underlying ggplot object
vpcdb	boolean whether to return the underlying vpcdb rather than the plot
verbose	show debugging information (TRUE or FALSE)

Value

a list containing calculated VPC information, and a ggplot2 object

See Also

vpc

---

vpc_tte *VPC function for survival-type data*

Description

Creates a VPC plot from observed and simulation survival data

Usage

```
vpc_tte(sim = NULL, obs = NULL, psn_folder = NULL, rtte = FALSE,
  rtte_calc_diff = TRUE, events = NULL, bins = FALSE, n_bins = 10,
  software = "auto", obs_cols = NULL, sim_cols = NULL, kmmc = NULL,
  reverse_prob = FALSE, stratify = NULL, stratify_color = NULL,
  ci = c(0.05, 0.95), plot = FALSE, xlab = NULL, ylab = NULL,
  show = NULL, as_percentage = TRUE, title = NULL, smooth = FALSE,
  vpc_theme = NULL, facet = "wrap", labeller = NULL, verbose = FALSE,
  vpcdb = FALSE)
```

Arguments

sim	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
obs	a data.frame with observed data, containing the indenpendent and dependent variable, a column indicating the individual, and possibly covariates. E.g. load in from NONMEM using read_table_nm
psn_folder	instead of specyfing "sim" and "obs", specify a PsN-generated VPC-folder
rtte	repeated time-to-event data? Deafult is FALSE (treat as single-event TTE)
rtte_calc_diff	recalculate time (T/F)? When simulating in NONMEM, you will probably need to set this to TRUE to recalculate the TIME to relative times between events (unless you output the time difference between events and specify that as independent variable to the vpc_tte() funciton.
events	numeric vector describing which events to show a VPC for when repeated TTE data, e.g. c(1:4). Default is NULL, which shows all events.
bins	either "density", "time", or "data", or a numeric vector specifying the bin separators.
n_bins	number of bins
software	name of software platform using (eg nonmem, phoenix)
obs_cols	observation dataset column names (list elements: "dv", "idv", "id", "pred")
sim_cols	simulation dataset column names (list elements: "dv", "idv", "id", "pred")
kmmc	either NULL (for regular TTE vpc, default), or a variable name for a KMMC plot (e.g. "WT")
reverse_prob	reverse the probability scale (i.e. plot 1-probability)
stratify	character vector of stratification variables. Only 1 or 2 stratification variables can be supplied. If stratify_color is also specified, only 1 additional stratification can be specified.
stratify_color	variable to stratify and color lines for observed data. Only 1 stratification variables can be supplied.
ci	confidence interval to plot. Default is (0.05, 0.95)
plot	Boolean indacting whether to plot the ggplot2 object after creation. Default is FALSE.
xlab	ylab as numeric vector of size 2
ylab	ylab as numeric vector of size 2
show	what to show in VPC (obs_ci, obs_median, sim_median, sim_median_ci)
as_percentage	Show y-scale from 0-100 percent? TRUE by default, if FALSE then scale from 0-1.
title	title

smooth	"smooth" the VPC (connect bin midpoints) or show bins as rectangular boxes. Default is TRUE.
vpc_theme	theme to be used in VPC. Expects list of class vpc_theme created with function vpc_theme()
facet	either "wrap", "columns", or "rows"
labeller	ggplot2 labeller function to be passed to underlying ggplot object
verbose	TRUE or FALSE (default)
vpcdb	Boolean whether to return the underlying vpcdb rather than the plot

Value

a list containing calculated VPC information, and a ggplot2 object

See Also

vpc

Examples

```
## Example for repeated) time-to-event data
## with NONMEM-like data (e.g. simulated using a dense grid)
data(rtte_obs_nm)
data(rtte_sim_nm)

# treat RTTE as TTE, no stratification
vpc_tte(sim = rtte_sim_nm,
        obs = rtte_obs_nm,
        rtte = FALSE,
        sim_cols=list(dv = "dv", idv = "t"), obs_cols=list(idv = "t"))

# stratified for covariate and study arm
vpc_tte(sim = rtte_sim_nm,
        obs = rtte_obs_nm,
        stratify = c("sex","drug"),
        rtte = FALSE,
        sim_cols=list(dv = "dv", idv = "t"), obs_cols=list(idv = "t"))

# stratified per event number (we'll only look at first 3 events) and stratify per arm
vpc_tte(sim = rtte_sim_nm,
        obs = rtte_obs_nm,
        rtte = TRUE, events = c(1:3),
        stratify = c("drug"),
        sim_cols=list(dv = "dv", idv = "t"), obs_cols=list(idv = "t"))
```