Package ‘arm’

August 1, 2007

Version 1.0-25

Date 2007-08-01

Title Data Analysis Using Regression and Multilevel/Hierarchical Models

Author Andrew Gelman <gelman@stat.columbia.edu>, Yu-Sung Su <ys463@columbia.edu>, Jennifer Hill <jh1030@columbia.edu>, Maria Grazia Pittau <grazia@stat.columbia.edu>, Jouni Kerman <jouni@kerman.com> and Tian Zheng <tzheng@stat.columbia.edu>

Maintainer Yu-Sung Su <ys463@columbia.edu>

Depends methods, R (>= 2.4.0), MASS, Matrix (>= 0.9975-1), lme4 (>= 0.9975-1), R2WinBUGS

Suggests car, foreign, nnet

Description R functions for processing lm, glm, mer and polr outputs.

URL http://www.stat.columbia.edu/~gelman/software/

License GPL (version 2 or later)

R topics documented:

`balanceplot` *Plot of Balance Statistics*

Description

This function plots the balance statistics before and after matching.

Usage

```
balanceplot(matched, pscore.fit,
    longcovnames = NULL,
    main = "Standardized Difference in Means",
    cex.main = 1, cex.vars = 0.8, cex.pts = 0.8,
    mar = c(0, 5, 4, 2), mgp = c(2, 0.25, 0),
    oma = c(0, 0, 0, 0), tcl = -0.2, ...)
```

Arguments

`matched`	matched data using `matching` function, see the example below.
`pscore.fit`	glm.fit object to get propensity scores.
`longcovnames`	long covariate names. If not provided, plot will use covariate variable name by default
`main`	title of the plot
`cex.main`	font size of main title
`cex.vars`	font size of variabel names
`cex.pts`	point size of the estimates
`mar`	margin of the plot, see ?`par` for details
`mgp`	axis margin of the plot, see ?`par` for details
`oma`	outer margin of the plot, see ?`par` for details
`tcl`	length of ticks, see ?`par` for details
`...`	other plot options may be passed to this function

Details

This function plots the balance statistics before and after matching. The open circle dots represent the unmatched balance statistics. The solid dots represent the matched balance statistics. The closer the value of the estimates to the zero, the better the treated and control groups are balanced after matching.

Note

The function does not work with predictors that contain factor(x), log(x) or all other data transformation. Create new objects for these variables. Attach them into the original dataset before doing the matching procedure.

Author(s)

Jennifer Hill ⟨jh1030@columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006. (Chater 10)

See Also

matching, par

Examples

```
data(lalonde)
attach(lalonde)
fit <- glm(treat ~ re74 + re75 + age + factor(educ) +
            black + hisp + married + nodegr + u74 + u75,
            family=binomial(link="logit"))
pscores <- predict(fit, type="link")
matches <- matching(z=lalonde$treat, score=pscores)
matched <- lalonde[matches$matched,]
balanceplot(matched, fit)
```

---

bayesglm *Bayesian generalized linear models.*

Description

Bayesian functions for generalized linear modeling with independent normal, t, or Cauchy prior distribution for the coefficients.

Usage

```
bayesglm (formula, family = gaussian, data,
    weights, subset, na.action,
    start = NULL, etastart, mustart,
    offset, control = glm.control(...),
    model = TRUE, method = "glm.fit",
    x = FALSE, y = TRUE, contrasts = NULL,
    prior.mean = 0, prior.scale = 2.5,
    prior.scale.for.intercept = 10,
    prior.df = 1, min.prior.scale=1e-12,
    scaled = TRUE, drop.baseline=TRUE, n.iter = 100, ...)

bayesglm.fit (x, y, weights = rep(1, nobs),
```

```
start = NULL, etastart = NULL,
mustart = NULL, offset = rep(0, nobs), family = gaussian(),
control = glm.control(), intercept = TRUE,
prior.mean = 0, prior.scale = 2.5,
prior.scale.for.intercept = 10,
prior.df = 1, min.prior.scale=1e-12, scaled = TRUE)
```

Arguments

`formula`	a symbolic description of the model to be fit. The details of model specification are given below.
`family`	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See `family` for details of family functions.)
`data`	an optional data frame, list or environment (or object coercible by `as.data.frame` to a data frame) containing the variables in the model. If not found in `data`, the variables are taken from `environment(formula)`, typically the environment from which `glm` is called.
`weights`	an optional vector of weights to be used in the fitting process. Should be `NULL` or a numeric vector.
`subset`	an optional vector specifying a subset of observations to be used in the fitting process.
`na.action`	a function which indicates what should happen when the data contain `NA`s. The default is set by the `na.action` setting of `options`, and is `na.fail` if that is unset. The "factory-fresh" default is `na.omit`. Another possible value is `NULL`, no action. Value `na.exclude` can be useful.
`start`	starting values for the parameters in the linear predictor.
`etastart`	starting values for the linear predictor.
`mustart`	starting values for the vector of means.
`offset`	this can be used to specify an *a priori* known component to be included in the linear predictor during fitting. This should be `NULL` or a numeric vector of length either one or equal to the number of cases. One or more `offset` terms can be included in the formula instead or as well, and if both are specified their sum is used. See `model.offset`.
`control`	a list of parameters for controlling the fitting process. See the documentation for `glm.control` for details.
`model`	a logical value indicating whether *model frame* should be included as a component of the returned value.
`method`	the method to be used in fitting the model. The default method `"glm.fit"` uses iteratively reweighted least squares (IWLS). The only current alternative is `"model.frame"` which returns the model frame and does no fitting.
`x, y`	For `glm`: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For `glm.fit`: `x` is a design matrix of dimension `n * p`, and `y` is a vector of observations of length `n`.
`contrasts`	an optional list. See the `contrasts.arg` of `model.matrix.default`.

`intercept`	logical. Should an intercept be included in the *null* model?
`prior.mean`	prior mean for the coefficients: default is 0. Can be a vector of length equal to the number of predictors (including the intercept, if any). If it is a scalar, it is expanded to the length of this vector.
`prior.scale`	prior scale for the coefficients: default is 2.5. Can be a vector of length equal to the number of predictors (including the intercept, if any). If it is a scalar, it is expanded to the length of this vector.
`prior.scale.for.intercept`	prior scale for the intercept: default is 10.
`prior.df`	for t distribution: default is 1 (Cauchy). Set to Inf to get normal prior distributions. Can be a vector of length equal to the number of predictors (including the intercept, if any). If it is a scalar, it is expanded to the length of this vector.
`min.prior.scale`	Minimum prior scale for the coefficients: default is 1e-12.
`scaled`	if scaled = TRUE, then the prior distribution is rescaled: default is TRUE
`drop.baseline`	Drop the base level of categorical x's, default is TRUE.
`n.iter`	default is 100.
`...`	further arguments passed to or from other methods.

Details

The program is a simple alteration of `glm()` that uses an approximate EM algorithm to update the betas at each step using an augmented regression to represent the prior information.

We use Student-t prior distributions for the coefficients. The prior distribution for the constant term is set so it applies to the value when all predictors are set to their mean values.

If scaled=TRUE, the scales for the prior distributions of the coefficients are determined as follows: For a predictor with only one value, we just use prior.scale. For a predictor with two values, we use prior.scale/range(x). For a predictor with more than two values, we use prior.scale/(2*sd(x)).

We include all the `glm()` arguments but we haven't tested that all the options (e.g., `offests`, `contrasts`, `deviance` for the null model) all work.

The new arguments here are: `prior.mean`, `prior.scale`, `prior.scale.for.intercept`, `prior.df`, and `scaled`.

Value

See `glm` for details.

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩; Maria Grazia Pittau ⟨grazia@stat.columbia.edu⟩; Aleks Jakulin ⟨Jakulin@stat.columbia.edu⟩

References

Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su, A default prior distribution for logistic and other regression models, unpublished paper available at http://www.stat.columbia.edu/~gelman/standardize/

See Also

glm, bayespolr

Examples

```
    n <- 100
    x1 <- rnorm (n)
    x2 <- rbinom (n, 1, .5)
    b0 <- 1
    b1 <- 1.5
    b2 <- 2
    y <- rbinom (n, 1, invlogit(b0+b1*x1+b2*x2))

    M1 <- glm (y ~ x1 + x2, family=binomial(link="logit"))
    display (M1)

    M2 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=Inf, prior.df=Inf)
    display (M2)  # just a test:  this should be identical to classical logit

    M3 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"))
      # default Cauchy prior with scale 2.5
    display (M3)

    M4 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=2.5, prior.df=1)
      # Same as M3, explicitly specifying Cauchy prior with scale 2.5
    display (M4)

    M5 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=2.5, prior.df=7)   # t_7 prior with scale 2.5
    display (M5)

    M6 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=2.5, prior.df=Inf)  # normal prior with scale 2.5
    display (M6)

  # Create separation:  set y=1 whenever x2=1
  # Now it should blow up without the prior!

    y <- ifelse (x2==1, 1, y)

    M1 <- glm (y ~ x1 + x2, family=binomial(link="logit"))
    display (M1)

    M2 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=Inf, prior.df=Inf) # Same as M1
    display (M2)

    M3 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"))
    display (M3)

    M4 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
      prior.scale=2.5, prior.df=1)  # Same as M3
    display (M4)
```

```
M5 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
  prior.scale=2.5, prior.df=7)
display (M5)

M6 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"),
  prior.scale=2.5, prior.df=Inf)
display (M6)

# bayesglm with gaussian family (bayes lm)
sigma <- 5
y2 <- rnorm (n, b0+b1*x1+b2*x2, sigma)
M7 <- bayesglm (y2 ~ x1 + x2, prior.scale=Inf, prior.df=Inf)
display (M7)


# bayesglm with categorical variables
z1 <- trunc(runif(n, 4, 9))
levels(factor(z1))
z2 <- trunc(runif(n, 15, 19))
levels(factor(z2))

## drop the base level (R default)
M8 <- bayesglm (y ~ x1 + factor(z1) + factor(z2),
  family=binomial(link="logit"), prior.scale=2.5, prior.df=Inf)
display (M8)

## keep all levels
M9 <- bayesglm (y ~ x1 + factor(z1) + factor(z2),
  family=binomial(link="logit"), prior.scale=2.5, prior.df=Inf,
  drop.baseline=FALSE)
display (M9)

M10 <- bayesglm (y ~ x1 + factor(z1) + factor(z2)-1,
  family=binomial(link="logit"), prior.scale=2.5, prior.df=Inf,
  drop.baseline=FALSE)
display (M10)
```

`bayespolr` *Bayesian Ordered Logistic or Probit Regression*

Description

Bayesian functions for ordered logistic or probit modeling with independent normal, t, or Cauchy prior distribution for the coefficients.

Usage

```
bayespolr(formula, data, weights, start, ...,
    subset, na.action, contrasts = NULL,
    Hess = TRUE, model = TRUE,
    method = c("logistic", "probit", "cloglog", "cauchit"),
    drop.unused.levels=TRUE,
    prior.mean = 0, prior.scale = 2.5, prior.df = 1,
```

```
    scaled = TRUE, prior.mean.for.cutpoints = 0,
    prior.scale.for.cutpoints = 10, prior.df.for.cutpoints = 1,
    n.iter = 100)
```

Arguments

`formula` a formula expression as for regression models, of the form 'response predictors'. The response should be a factor (preferably an ordered factor), which will be interpreted as an ordinal response, with levels ordered as in the factor. A proportional odds model will be fitted. The model must have an intercept: attempts to remove one will lead to a warning and be ignored. An offset may be used. See the documentation of 'formula' for other details.

`data` an optional data frame in which to interpret the variables occurring in 'formula'.

`weights` optional case weights in fitting. Default to 1.

`start` initial values for the parameters. This is in the format 'c(coefficients, zeta)'

`...` additional arguments to be passed to 'optim', most often a 'control' argument.

`subset` expression saying which subset of the rows of the data should be used in the fit. All observations are included by default.

`na.action` a function to filter missing data.

`contrasts` a list of contrasts to be used for some or all of the factors appearing as variables in the model formula.

`Hess` logical for whether the Hessian (the observed information matrix) should be returned.

`model` logical for whether the model matrix should be returned.

`method` logistic or probit or complementary log-log or cauchit (corresponding to a Cauchy latent variable and only available in R >= 2.1.0).

`drop.unused.levels`
default TRUE, if FALSE, it interpolates the intermediate values if the data have integer levels.

`prior.mean` prior mean for the coefficients: default is 0. Can be a vector of length equal to the number of predictors (not counting the intercepts). If it is a scalar, it is expanded to the length of this vector.

`prior.scale` prior scale for the coefficients: default is 2.5. Can be a vector of length equal to the number of predictors (not counting the intercepts). If it is a scalar, it is expanded to the length of this vector.

`prior.df` for t distribution: default is 1 (Cauchy). Set to Inf to get normal prior distributions. Can be a vector of length equal to the number of predictors (not counting the intercepts). If it is a scalar, it is expanded to the length of this vector.

`scaled` if scaled = TRUE, then the prior distribution is rescaled. Can be a vector of length equal to the number of cutpoints (intercepts). If it is a scalar, it is expanded to the length of this vector.

`prior.mean.for.cutpoints`
prior mean for cutpoints: default is 0. Can be a vector of length equal to the number of cutpoints (intercepts). If it is a scalar, it is expanded to the length of this vector.

`prior.scale.for.cutpoints`
prior scale for cutpoints: default is 10. Can be a vector of length equal to the number of cutpoints (intercepts). If it is a scalar, it is expanded to the length of this vector.

`prior.df.for.cutpoints`
: for t distribution: default is 1 (Cauchy). Can be a vector of length equal to the number of cutpoints (intercepts). If it is a scalar, it is expanded to the length of this vector.

`n.iter`
: default is 100.

Details

The program is a simple alteration of `polr` in `VR` version 7.2-31 that augments the loglikelihood with the log of the t prior distributions for the coefficients.

We use Student-t prior distributions for the coefficients. The prior distributions for the intercepts (the cutpoints) are set so they apply to the value when all predictors are set to their mean values.

If scaled=TRUE, the scales for the prior distributions of the coefficients are determined as follows: For a predictor with only one value, we just use `prior.scale`. For a predictor with two values, we use prior.scale/range(x). For a predictor with more than two values, we use prior.scale/(2*sd(x)).

Value

See `polr` for details.

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩; Maria Grazia Pittau ⟨grazia@stat.columbia.edu⟩

See Also

`bayesglm`, `polr`

Examples

```
M1 <- polr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
display (M1)

M2 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing,
    prior.scale=Inf, prior.df=Inf) # Same as M1
display (M2)

M3 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
display (M3)

M4 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing,
    prior.scale=2.5, prior.df=1)  # Same as M3
display (M4)

M5 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing,
    prior.scale=2.5, prior.df=7)
display (M5)

M6 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing,
    prior.scale=2.5, prior.df=Inf)
display (M6)
```

binnedplot *Binned Residual Plot*

Description

A function that plots averages of y versus averages of x and can be useful to plot residuals for logistic regression.

Usage

```
binnedplot(x ,y, nclass=floor(sqrt(length(x))),
   xlab="Expected Values", ylab="Average residual",
   main="Binned residual plot",
   cex.pts=0.8, col.pts=1, col.int="gray")
```

Arguments

x	The expected values from the logistic regression.
y	The residuals values from logistic regression (observed values minus expected values).
nclass	Number of categories (bins) based on their fitted values in which the data are divided, default is floor(sqrt(length(x))).
xlab	a label for the x axis, default is "Expected Values".
ylab	a label for the y axis, default is "Average residual".
main	a main title for the plot, default is "Binned residual plot". See also `title`.
cex.pts	The size of points, default=0.8.
col.pts	color of points, default is black
col.int	color of intervals, default is gray

Details

In logistic regression, as with linear regression, the residuals can be defined as observed minus expected values. The data are discrete and so are the residuals. As a result, plots of raw residuals from logistic regression are generally not useful. The binned residuals plot instead, after dividing the data into categories (bins) based on their fitted values, plots the average residual versus the average fitted value for each bin.

Value

A plot in which the gray lines indicate ± 2 standard-error bounds, within which one would expect about 95% of the binned residuals to fall, if the model were actually true.

Note

There is typically some arbitrariness in choosing the number of bins: each bin should contain enough points so that the averaged residuals are not too noisy, but it helps to have also many bins so as to see more local patterns in the residuals (see Gelman and Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, pag 97).

Author(s)

M. Grazia Pittau ⟨grazia@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

par, plot

Examples

```
data(lalonde)
attach(lalonde)
fit <- glm(treat ~ re74 + re75 + educ + black + hisp + married
              + nodegr + u74 + u75, family=binomial(link="logit"))
x <- predict(fit)
y <- resid(fit)
binnedplot(x,y)
```

`coefplot` *Generic Function for Making Coefficient Plot*

Description

Functions that plot the coefficients $\pm$ 1 and 2 sd from a lm, glm, bugs, and polr fits.

Usage

```
coefplot(object, ...)

coefplot.default(coefs, sds,
                varnames=NULL, CI=2, vertical=TRUE,
                xlim=NULL, ylim=NULL,
                xlab="", ylab="", main="",
                cex.var=0.8, cex.pts=0.9, col.pts=1,
                var.las=2)

# methods for coefplot()
coefplot.lm (object, varnames=NULL, intercept=FALSE, ...)
coefplot.glm (object, varnames=NULL, intercept=FALSE, ...)
coefplot.polr (object, varnames=NULL, ...)
coefplot.bugs (object, varnames=NULL, CI=2,
               xlim=NULL, ylim=NULL,
               xlab="", ylab="", main="",
               cex.var=0.8, cex.pts=0.9, col.pts=1)
```

Arguments

`object`	fitted objects-lm, glm, bugs and polr, or a vector of coefficients.
`...`	further arguments passed to or from other methods.
`coefs`	a vector of coefficients.
`sds`	a vector of sds of coefficients.
`varnames`	a vector of variable names, default is NULL, which will use the names of variables.
`CI`	confidence interval, default is 2, which will plot ± 2 sds or 95% CI. If CI=1, plot ± 1 sds or 50% CI instead.
`vertical`	orientation of the plot, default is TRUE which will plot variable names in the 2nd axis. If FALSE, plot variable names in the first axis instead.
`xlim`	the x limits (x1, x2) of the plot. Note that 'x1 > x2' is allowed and leads to a "reversed axis".
`ylim`	the y limits of the plot.
`xlab`	a label for the x axis, default is "".
`ylab`	a label for the y axis, default is "".
`main`	a main title for the plot, default is "". See also `title`.
`cex.var`	The fontsize of the varible names, default=0.8.
`cex.pts`	The size of data points, default=0.9.
`col.pts`	color of points and segments, default is black.
`var.las`	the orientation of variable names against the axis, default is 2. see the usage of `las` in `par`.
`intercept`	If TRUE will plot intercept, default=FALSE to get better presentation.

Details

This function plots coefficients from lm, glm and polr with 1 sd and 2 sd interval bars.

Value

Plot of the coefficients from a lm or glm fit. You can add the intercept, the variable names and the display the result of the fitted model.

Author(s)

Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

`display`, `par`, `lm`, `glm`, `bayesglm`

Examples

```
y1 <- rnorm(1000,50,23)
y2 <- rbinom(1000,1,prob=0.72)
x1 <- rnorm(1000,50,2)
x2 <- rbinom(1000,1,prob=0.63)
x3 <- rpois(1000, 2)
x4 <- runif(1000,40,100)
x5 <- rbeta(1000,2,2)

longnames <- c("a long name01","a long name02","a long name03",
               "a long name04","a long name05")

fit1 <- lm(y1 ~ x1 + x2 + x3 + x4 + x5)
fit2 <- glm(y2 ~ x1 + x2 + x3 + x4 + x5,
           family=binomial(link="logit"))

# plot 1
par (mfrow=c(2,2))
coefplot(fit1)
coefplot(fit2, col.pts="blue")


# plot 2
par (mar=c(2,8,2,0.5))
coefplot(fit1, longnames, intercept=TRUE, CI=1)

# plot 3
par (mar=c(2,2,2,2))
coefplot(fit2, vertical=FALSE, var.las=1)

# plot 4: comparison to show bayesglm works better than glm
n <- 100
x1 <- rnorm (n)
x2 <- rbinom (n, 1, .5)
b0 <- 1
b1 <- 1.5
b2 <- 2
y <- rbinom (n, 1, invlogit(b0+b1*x1+b2*x2))
y <- ifelse (x2==1, 1, y)
x1 <- rescale(x1)
x2 <- rescale(x2, "center")

M1 <- glm (y ~ x1 + x2, family=binomial(link="logit"))
      display (M1)
M2 <- bayesglm (y ~ x1 + x2, family=binomial(link="logit"))
      display (M2)

   ## stacked plot
   par(mar=c(2,5,3,1), mgp=c(2,0.25,0), oma=c(0,0,2,0), tcl=-0.2)

   coefplot(M2, xlim=c(-1,5), intercept=TRUE)
   points(coef(M1), c(3:1)-0.1, col="red", pch=19)
   segments(coef(M1) + se.coef(M1), c(3:1)-0.1,
       coef(M1) - se.coef(M1), c(3:1)-0.1, lwd=2, col="red")
   segments(coef(M1) + 2*se.coef(M1), c(3:1)-0.1,
       coef(M1) - 2*se.coef(M1), c(3:1)-0.1, col="red")
```

```
    mtext("Coefficients", side=3, at=0.1, outer=TRUE)
    mtext("Estimate", side=3, at=0.6, outer=TRUE)

    ## arrayed plot
    par(mfrow=c(1,2), mar=c(2,5,5,1), mgp=c(2,0.25,0), tcl=-0.2)
    x.scale <- c(0, 7.5) # fix x.scale for comparison

    coefplot(M1, xlim=x.scale, main="glm", intercept=TRUE)
    coefplot(M2, xlim=x.scale, main="bayesglm", intercept=TRUE)

# plot 5: the ordered logit model from polr
 M3 <- polr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
 par (mar=c(2,8,2,0.5))
 coefplot(M3)

 M4 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
 par (mar=c(2,8,2,0.5))
 coefplot(M4)

# plot 6: plot bugs
 M5 <- lmer(Reaction ~ Days + (1|Subject), sleepstudy)
 M5.sim <- mcsamp(M5)
 coefplot(M5.sim)

# plot 7: plot coefficients & sds vectors
 coef.vect <- c(0.2, 1.4, 2.3, 0.5)
 sd.vect <- c(0.12, 0.24, 0.23, 0.15)
 longnames <- c("var1", "var2", "var3", "var4")
 coefplot (coef.vect, sd.vect, longnames)
 coefplot (coef.vect, sd.vect, longnames, vertical=FALSE, var.las=1)
```

`contrast.bayes.ordered` *Contrast Matrices*

Description

Return a matrix of contrasts used in `bayesglm`.

Usage

```
contr.bayes.unordered(n, base = 1, contrasts = TRUE)
contr.bayes.ordered (n, scores = 1:n, contrasts = TRUE)
```

Arguments

`n`	a vector of levels for a factor, or the number of levels.
`base`	an integer specifying which group is considered the baseline group. Ignored if `contrasts` is `FALSE`.
`contrasts`	a logical indicating whether contrasts should be computed.
`scores`	the set of values over which orthogonal polynomials are to be computed.

Details

These functions are adapted from `contr.treatment` and `contr.poly` in `stats` package. The purpose for these functions are to keep the baseline levels of categorical variables and thus to suit the use of `bayesglm`.

`contr.bayes.unordered` is equivalent to `contr.treatment` whereas `contr.bayes.ordered` is equivalent to `contr.poly`.

Author(s)

Yu-Sung Su ⟨ys463@columbia.edu⟩

See Also

`C`, `contr.helmert`, `contr.poly`, `contr.sum`, `contr.treatment`; `glm`, `aov`, `lm`, `bayesglm`.

Examples

```
cat.var <- rep(1:3, 5)
dim(contr.bayes.unordered(cat.var))
# 15*15 baseline level kept!
dim(contr.treatment(cat.var))
# 15*14
```

`corrplot` *Correlation Plot*

Description

Function for making a correlation plot starting from a data matrix

Usage

```
corrplot (data, varnames=NULL, cutpts=NULL,
    abs=TRUE, details=TRUE,
    n.col.legend=5, cex.col=0.7,
    cex.var=0.9, digits=1, color=FALSE)
```

Arguments

`data`	a data matrix
`varnames`	variable names of the data matrix, if not provided use default variable names
`abs`	if TRUE, transform all correlation values into positive values, default=TRUE.
`cutpts`	a vector of cutting points for color legend, default is NULL. The function will decide the cutting points if cutpts is not assigned.
`details`	show more than one digits correlaton values. Default is TRUE. FALSE is suggested to get readable output.
`n.col.legend`	number of legend for the color thermometer.
`cex.col`	font size of the color thermometer.
`cex.var`	font size of the variable names.
`digits`	number of digits shown in the text of the color theromoeter.
`color`	color of the plot, default is FALSE, which uses gray scale.

Details

The function adapts the R function for Figure 8 in Tian Zheng, Matthew Salganik, and Andrew Gelman, 2006, "How many people do you know in prison?: using overdispersion in count data to estimate social structure in networks", Journal of the American Statistical Association, Vol.101, N0. 474: p.409-23.

Value

A correlation plot.

Note

see also: http://www.stat.columbia.edu/~gelman/research/published/

Author(s)

Tian Zheng ⟨tzheng@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Tian Zheng, Matthew Salganik, and Andrew Gelman, 2006, "How many people do you know in prison?: using overdispersion in count data to estimate social structure in networks", Journal of the American Statistical Association, Vol.101, N0. 474: p.409-23

See Also

cor, par

Examples

```
x1 <- rnorm(1000,50,2)
x2 <- rbinom(1000,1,prob=0.63)
x3 <- rpois(1000, 2)
x4 <- runif(1000,40,100)
x5 <- rnorm(1000,100,30)
x6 <- rbeta(1000,2,2)
x7 <- rpois(1000,10)
x8 <- rbinom(1000,1,prob=0.4)
x9 <- rbeta(1000,5,4)
x10 <- runif(1000,-10,-1)

test.data <- data.matrix(cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10))
test.names <- c("a short name01","a short name02","a short name03",
                "a short name04","a short name05","a short name06",
                "a short name07","a short name08","a short name09",
                "a short name10")

# example 1
corrplot(test.data)

# example 2
corrplot(test.data,test.names, abs=FALSE, n.col.legend=7)
corrplot(test.data,test.names, abs=TRUE, n.col.legend=7)

# example 3
```

```
data(lalonde)
corrplot(lalonde, details=FALSE, color=TRUE)
corrplot(lalonde, cutpts=c(0,0.25,0.5,0.75), color=TRUE, digits=2)
```

---

display *Functions for Processing lm, glm, mer and polr Output*

Description

This generic function gives a clean printout of lm, glm, mer and polr objects.

Usage

```
    display (object, ...)

# methods for display()
    display.lm (object, digits=2)
    display.glm (object, digits=2)
    display.mer (object, digits=2)
    display.mer2 (object, digits=2)
    display.polr (object, digits=2)
```

Arguments

`object`	The output of a call to lm, glm, mer, polr, or related regressions function with n data points and k predictors.
`...`	further arguments passed to or from other methods.
`digits`	number of significant digits to display.

Details

This generic function gives a clean printout of lm, glm, mer and polr objects, focusing on the most pertinent pieces of information: the coefficients and their standard errors, the sample size, number of predictors, residual standard deviation, and R-squared. Note: R-squared is automatically displayed to 2 digits, and deviances are automatically displayed to 1 digit, no matter what.

Value

Coefficients and their standard errors, the sample size, number of predictors, residual standard deviation, and R-squared

Note

Output are the model, the regression coefficients and standard errors, and the residual sd and R-squared (for a linear model), or the null deviance and residual deviance (for a generalized linear model).

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩; Maria Grazia Pittau ⟨grazia@stat.columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

summary, lm, glm, lmer, polr

Examples

```
# Here's a simple example of a model of the form, y = a + bx + error,
# with 10 observations in each of 10 groups, and with both the
# intercept and the slope varying by group.  First we set up the model and data.

   group <- rep(1:10, rep(10,10))
   mu.a <- 0
   sigma.a <- 2
   mu.b <- 3
   sigma.b <- 4
   rho <- 0
   Sigma.ab <- array (c(sigma.a^2, rho*sigma.a*sigma.b,
                    rho*sigma.a*sigma.b, sigma.b^2), c(2,2))
   sigma.y <- 1
   ab <- mvrnorm (10, c(mu.a,mu.b), Sigma.ab)
   a <- ab[,1]
   b <- ab[,2]
#
   x <- rnorm (100)
   y1 <- rnorm (100, a[group] + b[group]*x, sigma.y)
   y2 <- rbinom(100, 1, prob=invlogit(a[group] + b*x))

#
# display a simple linear model
#
   M1 <- lm (y1 ~ x)
   display (M1)

#
# display a simple logit model
#
   M2 <- glm (y2 ~ x, family=binomial(link="logit"))
   display (M2)
#
# Then fit and display a simple varying-intercept model:
#
   M3 <- lmer (y1 ~ x + (1|group))
   display (M3)
   M3.sim <- mcsamp (M3)
   print (M3.sim)
   plot (M3.sim)
#
# Then the full varying-intercept, varying-slope model:
#
   M4 <- lmer (y1 ~ x + (1 + x |group))
   display (M4)
   M4.sim <- mcsamp (M4)
```

```
   print (M4.sim)
   plot (M4.sim)
#
# Then the full varying-intercept, logit model:
#
   M5 <- lmer (y2 ~ x + (1|group), family=binomial(link="logit"))
   display (M5)
   M5.sim <- mcsamp (M5)
   print (M5.sim)
   plot (M5.sim)
#
# Then the full varying-intercept, varying-slope logit model:
#
   M6 <- lmer (y2 ~ x + (1 + x |group), family=binomial(link="logit"))
   display (M6)
   M6.sim <- mcsamp (M6)
   print (M6.sim)
   plot (M6.sim)

#
# Then the ordered logit model from polr
#
   M7 <- polr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
   display(M7)

   M8 <- bayespolr(Sat ~ Infl + Type + Cont, weights = Freq, data = housing)
   display(M8)
```

`fround` *Formating the Rounding of Numbers*

Description

`fround` rounds the values in its first argument to the specified number of decimal places with surrounding quotes.

`pfround` rounds the values in its first argument to the specified number of decimal places without surrounding quotes.

Usage

```
fround(x, digits)
pfround(x, digits)
```

Arguments

`x`	a numeric vector.
`digits`	integer indicating the precision to be used.

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

See Also

round

Examples

```
x <- rnorm(1)
fround(x, digits=2)
pfround(x, digits=2)
```

---

go *Function to Recall Last Source File*

Description

A function that like source() but recalls the last source file names by default.

Usage

```
go(..., add=FALSE, timer=FALSE)
```

Arguments

...	list of filenames as character strings.
add	add these names to the current list? if replace, then FALSE.
timer	time the execution time of go().

Author(s)

Jouni Kerman ⟨jouni@kerman.com⟩ ⟨kerman@stat.columbia.edu⟩

Examples

```
go('myprog')              # will run source('myprog.r')
go()                      # will run source('myprog.r') again
go('somelib',add=TRUE)    # will run source('myprog.r') and source('somelib.r')
go('myprog','somelib')    # same as above
go('mytest')              # will run source('mytest') only
go()                      # runs source('mytest') again
G                         # short cut to call go()
```

`invlogit` *Inverse logistic function*

Description

Inverse-logit function, transforms continuous values to the range (0, 1)

Usage

```
invlogit(x)
```

Arguments

`x`	A vector of continuous values

Details

The Inverse-logit function defined as: $logit^-1(x) = e^x/(1 + e^x)$ transforms continuous values to the range (0, 1), which is necessary, since probabilities must be between 0 and 1 and maps from the linear predictor to the probabilities

Value

A vector of estimated probabilities

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩, M.Grazia Pittau ⟨grazia@stat.columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

Examples

```
data(frisk)
n <- 100
x1 <- rnorm (n)
x2 <- rbinom (n, 1, .5)
b0 <- 1
b1 <- 1.5
b2 <- 2
Inv.logit <- invlogit(b0+b1*x1+b2*x2)
plot(b0+b1*x1+b2*x2, Inv.logit)
```

`lalonde` *Lalonde Dataset*

Description

Dataset used by Dehejia and Wahba (1999) to evaluate propensity score matching.

Usage

```
data(lalonde)
```

Format

A data frame with 445 observations on the following 12 variables.

age age in years.

educ years of schooling.

black indicator variable for blacks.

hisp indicator variable for Hispanics.

married indicator variable for martial status.

nodegr indicator variable for high school diploma.

re74 real earnings in 1974.

re75 real earnings in 1975.

re78 real earnings in 1978.

u74 indicator variable for earnings in 1974 being zero.

u75 indicator variable for earnings in 1975 being zero.

treat an indicator variable for treatment status.

Details

Two demos are provided which use this dataset. The first, `DehejiaWahba`, replicates one of the models from Dehejia and Wahba (1999). The second demo, `AbadieImbens`, replicates the models produced by Abadie and Imbens http://elsa.berkeley.edu/~imbens/matlab/lalonde_exper_04feb2.m. Many of these models are found to produce good balance for the Lalonde data.

Note

This documentation is adapted from `Matching` package.

References

Dehejia, Rajeev and Sadek Wahba. 1999."Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94 (448): 1053-1062.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs." *American Economic Review* 76:604-620.

See Also

matching, GenMatch balanceplot

Examples

```
data(lalonde)
```

---

matching *Matching*

Description

Function for processing matching with propensity score

Usage

```
matching(z, score)
```

Arguments

z	vector of indicators for treatment or control.
score	vector of the propensity scores in the same order as z.

Details

Function for matching each treatment unit in turn the control unit (not previously chosen) with the closest propensity score

Value

The function returns a vector of indices that the corresponding unit is matched to. 0 means matched to nothing.

Author(s)

Jeniffer Hill ⟨jh1030@columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

balanceplot

Examples

```
data(lalonde)
attach(lalonde)
fit <- glm(treat ~ re74 + re75 + age + factor(educ) +
            black + hisp + married + nodegr + u74 + u75,
            family=binomial(link="logit"))
pscores <- predict(fit, type="link")
matches <- matching(z=lalonde$treat, score=pscores)
matched <- lalonde[matches$matched,]
balanceplot(matched, fit)
```

---

mcsamp — *Quick Function to Run mcmcsamp() in lme4*

Description

The quick function for MCMC sampling for lmer and glmer objects and convert to Bugs objects for easy display.

Usage

```
mcsamp (object, n.chains=3, n.iter=1000,
    n.burnin=floor(n.iter/2), n.thin=1,
    saveb=TRUE, make.bugs.object=TRUE)
```

Arguments

`object`	lmer or glmer objects
`n.chains`	number of MCMC chains
`n.iter`	number of iteration for each MCMC chain
`n.burnin`	number of burnin for each MCMC chain
`n.thin`	number of thin for each MCMC chain
`saveb`	if 'TRUE', causes the values of the random effects in each sample to be saved.
`make.bugs.object`	tranform the output into bugs object, default is TRUE

Details

This function generates a sample from the posterior distribution of the parameters of a fitted model using Markov Chain Monte Carlo methods. It automatically simulates multiple sequences and allows convergence to be monitored. The function relies on `mcmcsamp` in `lme4`.

Value

An object of (S3) class '"bugs"' suitable for use with the functions in the "R2WinBUGS" package.

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

Douglas Bates and Deepayan Sarkar, lme4: Linear mixed-effects models using S4 classes.

See Also

`display`, `lmer`, `mcmcsamp`, `sim`

Examples

```
# Here's a simple example of a model of the form, y = a + bx + error,
# with 10 observations in each of 10 groups, and with both the intercept
# and the slope varying by group.  First we set up the model and data.
#
   group <- rep(1:10, rep(10,10))
   mu.a <- 0
   sigma.a <- 2
   mu.b <- 3
   sigma.b <- 4
   rho <- 0
   Sigma.ab <- array (c(sigma.a^2, rho*sigma.a*sigma.b,
            rho*sigma.a*sigma.b, sigma.b^2), c(2,2))
   sigma.y <- 1
   ab <- mvrnorm (10, c(mu.a,mu.b), Sigma.ab)
   a <- ab[,1]
   b <- ab[,2]
#
   x <- rnorm (100)
   y1 <- rnorm (100, a[group] + b[group]*x, sigma.y)
   y2 <- rbinom(100, 1, prob=invlogit(a[group] + b*x))
#
# Then fit and display a simple varying-intercept model:

   M1 <- lmer (y1 ~ x + (1|group))
   display (M1)
   M1.sim <- mcsamp (M1)
   print (M1.sim)
   plot (M1.sim)
#
# Then the full varying-intercept, varying-slope model:
#
   M2 <- lmer (y1 ~ x + (1 + x |group))
   display (M2)
   M2.sim <- mcsamp (M2)
   print (M2.sim)
   plot (M2.sim)
#
# Then the full varying-intercept, logit model:
#
   M3 <- lmer (y2 ~ x + (1|group), family=binomial(link="logit"))
   display (M3)
   M3.sim <- mcsamp (M3)
   print (M3.sim)
   plot (M3.sim)
#
```

```
# Then the full varying-intercept, varying-slope logit model:
#
   M4 <- lmer (y2 ~ x + (1 + x |group), family=binomial(link="logit"))
   display (M4)
   M4.sim <- mcsamp (M4)
   print (M4.sim)
   plot (M4.sim)
```

`model.matrix.bayes` *Construct Design Matrices*

Description

`model.matrix` creates a design matrix.

Usage

```
# bayesglm
model.matrix.bayes(object, data = environment(object),
    contrasts.arg = NULL, xlev = NULL, keep.order = FALSE, ...)

# bayesglm.hiearchical (not implement yet!)
model.matrix.bayes2(object, data = environment(object),
    contrasts.arg = NULL, xlev = NULL, keep.order = FALSE, batch = NULL, ...)
```

Arguments

`object` an object of an appropriate class. For the default method, a model formula or terms object.

`data` a data frame created with `model.frame`. If another sort of object, `model.frame` is called first.

`contrasts.arg` A list, whose entries are contrasts suitable for input to the `contrasts` replacement function and whose names are the names of columns of `data` containing `factors`.

`xlev` to be used as argument of `model.frame` if `data` has no `"terms"` attribute.

`keep.order` a logical value indicating whether the terms should keep their positions. If `FALSE` the terms are reordered so that main effects come first, followed by the interactions, all second-order, all third-order and so on. Effects of a given order are kept in the order specified.

`batch` Not implement yet!

`...` further arguments passed to or from other methods.

Details

`model.matrix.bayes` is adapted from `model.matrix` in the `stats` pacakge and is designed for the use of `bayesglm` and `bayesglm.hierachical` (not yet implemented!). It is designed to keep baseline levels of all categorical varaibles and keep the variable names unodered in the output. The design matrices created by `model.matrix.bayes` are unidentifiable using classical regression methods, though; they can be identified using `bayesglm` and `bayesglm.hierachical`.

Author(s)

Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman, Aleks Jakulin, Maria Grazia Pittau and Yu-Sung Su, A default prior distribution for logistic and other regression models, unpublished paper available at http://www.stat.columbia.edu/~gelman/standardize/

See Also

model.frame, model.extract, terms, terms.bayes, bayesglm.

Examples

```
ff <- log(Volume) ~ log(Height) + log(Girth)
str(m <- model.frame(ff, trees))
(model.matrix.bayes(ff, m))
(model.matrix.bayes2(ff, m))
(model.matrix(ff, m))
```

---

rescale *Function for Standardizing by Centering and Dividing by 2 sd's*

Description

This function standardizes a variable by centering and dividing by 2 sd's with exceptions for binary variables.

Usage

```
rescale(x, binary.inputs)
```

Arguments

x	a vector
binary.inputs	options for standardizing binary variables

Value

the standardized vector

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman, Scaling regression inputs by dividing by two standard deviations. http://www.stat.columbia.edu/~gelman/research/unpublished/standardizing.pdf

See Also

standardize

Examples

```
# Set up the fake data
n <- 100
x <- rnorm (n, 2, 1)
x1 <- rnorm (n)
x1 <- (x1-mean(x1))/(2*sd(x1))   # standardization
x2 <- rbinom (n, 1, .5)
b0 <- 1
b1 <- 1.5
b2 <- 2
y <- rbinom (n, 1, invlogit(b0+b1*x1+b2*x2))
rescale(x, "full2")
rescale(y, "center")
```

---

se.coef   *Extract Standard Errors of Model Coefficients*

Description

These functions extract standard errors of model coefficients from objects returned by modeling functions.

Usage

```
se.coef (object)
se.fixef (object)
se.ranef (object)

# methods for se.coef()
se.coef.lm (object)
se.coef.glm (object)
se.coef.mer (object)
```

Arguments

object	object of lm, glm, lmer and glmer fit

Details

`se.coef` extracts standard errors from objects returned by modeling functions. `se.fixef` extracts standard errors of the fixed effects from objects returned by lmer and glmer functions. `se.ranef` extracts standard errors of the random effects from objects returned by lmer and glmer functions.

Value

`se.coef` gives lists of standard errors for `coef`, `se.fixef` gives a vector of standard errors for `fixef` and `se.ranef` gives a list of standard errors for `ranef`.

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

display, coef, sigma.hat,

Examples

```
# Here's a simple example of a model of the form, y = a + bx + error,
# with 10 observations in each of 10 groups, and with both the
# intercept and the slope varying by group.  First we set up the model and data.

   group <- rep(1:10, rep(10,10))
   mu.a <- 0
   sigma.a <- 2
   mu.b <- 3
   sigma.b <- 4
   rho <- 0
   Sigma.ab <- array (c(sigma.a^2, rho*sigma.a*sigma.b,
                    rho*sigma.a*sigma.b, sigma.b^2), c(2,2))
   sigma.y <- 1
   ab <- mvrnorm (10, c(mu.a,mu.b), Sigma.ab)
   a <- ab[,1]
   b <- ab[,2]
#
   x <- rnorm (100)
   y1 <- rnorm (100, a[group] + b[group]*x, sigma.y)
   y2 <- rbinom(100, 1, prob=invlogit(a[group] + b*x))

#  lm fit
   M1 <- lm (y1 ~ x)
   se.coef (M1)

#  glm fit
   M2 <- glm (y2 ~ x)
   se.coef (M2)

#  lmer fit
   M3 <- lmer (y1 ~ x + (1 + x |group))
   se.coef (M3)
   se.fixef (M3)
   se.ranef (M3)

#  glmer fit
   M4 <- lmer (y2 ~ 1 + (0 + x |group), family=binomial(link="logit"))
   se.coef (M4)
   se.fixef (M4)
   se.ranef (M4)
```

`sigma.hat` *Extract Residual Errors*

Description

This generic function extracts residual errors from a fitted model.

Usage

```
sigma.hat(object)

# methods for sigma.hat()
sigma.hat.lm (object)
sigma.hat.glm (object)
sigma.hat.mer (object)
```

Arguments

`object`	any fitted model object of lm, glm and mer class

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩

See Also

`display`, `summary`, `lm`, `glm`, `lmer`

Examples

```
group <- rep(1:10, rep(10,10))
mu.a <- 0
sigma.a <- 2
mu.b <- 3
sigma.b <- 4
rho <- 0
Sigma.ab <- array (c(sigma.a^2, rho*sigma.a*sigma.b,
                 rho*sigma.a*sigma.b, sigma.b^2), c(2,2))
sigma.y <- 1
ab <- mvrnorm (10, c(mu.a,mu.b), Sigma.ab)
a <- ab[,1]
b <- ab[,2]

x <- rnorm (100)
y1 <- rnorm (100, a[group] + b[group]*x, sigma.y)
y2 <- rbinom(100, 1, prob=invlogit(a[group] + b*x))

M1 <- lm (y1 ~ x)
sigma.hat(M1)

M2 <- bayesglm (y1 ~ x, prior.scale=Inf, prior.df=Inf)
sigma.hat(M2) # should be same to sigma.hat(M1)

M3 <- glm (y2 ~ x, family=binomial(link="logit"))
```

```
sigma.hat(M3)

M4 <- lmer (y1 ~ (1+x|group))
sigma.hat(M4)

M5 <- lmer (y2 ~ (1+x|group), family=binomial(link="logit"))
sigma.hat(M5)
```

---

sim *Functions to Get Posterior Distributions*

Description

This generic function gets posterior simulations of sigma and beta from a lm object, or simulations of beta from a glm object, or simulations of beta from a mer object

Usage

```
sim(object, ...)

# methods for sim()
sim.lm (object, n.sims = 100)
sim.glm (object, n.sims = 100)
sim.mer (object, n.sims = 100)
```

Arguments

`object`	the output of a call to "lm" with n data points and k predictors.
`...`	further arguments passed to or from other methods.
`n.sims`	number of independent simulation draws to create.

Value

`sigma.sim`	vector of n.sims random draws of sigma (for glm's, this just returns a vector of 1's or else of the square root of the overdispersion parameter if that is in the model)
`beta.sim`	matrix (dimensions n.sims x k) of n.sims random draws of beta

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩; Yu-Sung Su ⟨ys463@columbia.edu⟩; M.Grazia Pittau ⟨grazia@stat.columbia.edu⟩

References

Andrew Gelman and Jennifer Hill, Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press, 2006.

See Also

`display`, `mcsamp`, `lm`, `glm`, `lmer`

Examples

```
#Examples of "sim"
 set.seed (1)
 J <- 15
 n <- J*(J+1)/2
 group <- rep (1:J, 1:J)
 mu.a <- 5
 sigma.a <- 2
 a <- rnorm (J, mu.a, sigma.a)
 b <- -3
 x <- rnorm (n, 2, 1)
 sigma.y <- 6
 y <- rnorm (n, a[group] + b*x, sigma.y)
 u <- runif (J, 0, 3)
 y123.dat <- as.data.frame (round (cbind (y, x, group), 2))

# Linear regression
 x1 <- y123.dat$y
 y1 <- y123.dat$x
 M1 <- lm (y1 ~ x1)
 display(M1)
 M1.sim <- sim(M1)

# Logistic regression
 u.data <- cbind (1:J, u)
 dimnames(u.data)[[2]] <- c("group", "u")
 u.dat <- as.data.frame (round (u.data, 2))
 y <- rbinom (n, 1, invlogit (a[group] + b*x))
 M2 <- glm (y ~ x, family=binomial(link="logit"))
 display(M2)
 M2.sim <- sim (M2)

# Using lmer:
# Example 1
 E1 <- lmer (y ~ x + (1 | group))
 display(E1)
 E1.sim <- sim (E1)

# Example 2
 u.full <- u[group]
 E2 <- lmer (y ~ x + u.full + (1 | group))
 display(E2)
 E2.sim <- sim (E2)

# Example 3
 y <- rbinom (n, 1, invlogit (a[group] + b*x))
 E3 <- lmer (y ~ x + (1 | group), family=binomial(link="logit"),
    control=list(usePQL=TRUE))
 display(E3)
 E3.sim <- sim (E3)
```

`standardize` *Function for Standardizing Regression Predictors by Centering and Dividing by 2 sd's*

Description

Numeric variables that take on more than two values are each rescaled to have a mean of 0 and a sd of 0.5; Binary variables are rescaled to have a mean of 0 and a difference of 1 between their two categories; Non-numeric variables that take on more than two values are unchanged; Variables that take on only one value are unchanged

Usage

```
standardize(object, unchanged = NULL,
    standardize.y = FALSE, binary.inputs = "center")
```

Arguments

`object`	an object of class "lm" or "glm"
`unchanged`	vector of names of parameters to leave unstandardized
`standardize.y`	if TRUE, the outcome variable is standardized also
`binary.inputs`	options for standardizing binary variables

Details

"0/1" (rescale so that the lower value is 0 and the upper is 1) "-0.5/0.5" (rescale so that the lower value is -0.5 and upper is 0.5) "center" (rescale so that the mean of the data is 0 and the difference between the two categories is 1) "full" (rescale by subtracting the mean and dividing by 2 sd's) "leave.alone" (do nothing)

Author(s)

Andrew Gelman ⟨gelman@stat.columbia.edu⟩ Yu-Sung Su ⟨ys463@columbia.edu⟩

References

Andrew Gelman, Scaling regression inputs by dividing by two standard deviations http://www.stat.columbia.edu/~gelman/research/unpublished/standardizing.pdf

See Also

rescale

Examples

```
# Set up the fake data
n <- 100
x <- rnorm (n, 2, 1)
x1 <- rnorm (n)
x1 <- (x1-mean(x1))/(2*sd(x1))   # standardization
x2 <- rbinom (n, 1, .5)
b0 <- 1
b1 <- 1.5
b2 <- 2
y <- rbinom (n, 1, invlogit(b0+b1*x1+b2*x2))
M1 <- glm (y ~ x, family=binomial(link="logit"))
display(M1)
```

```
M2 <-standardize(M1)
display(M2)
```

`terms.bayes` *Construct a terms Object from a Formula*

Description

This function takes a formula and some optional arguments and constructs a terms object. The terms object can then be used to construct a `model.matrix.bayes`.

Usage

```
terms.bayes(x, specials = NULL, abb = NULL, data = NULL,
    neg.out = TRUE, keep.order = FALSE, simplify = FALSE,
    allowDotAsName = FALSE, ...)
```

Arguments

`x`	a formula.
`specials`	which functions in the formula should be marked as special in the `terms` object.
`abb`	Not implemented in R.
`data`	a data frame from which the meaning of the special symbol `.` can be inferred. It is unused if there is no `.` in the formula.
`neg.out`	Not implemented in R.
`keep.order`	a logical value indicating whether the terms should keep their positions. If `FALSE` the terms are reordered so that main effects come first, followed by the interactions, all second-order, all third-order and so on. Effects of a given order are kept in the order specified. Default is `FALSE`.
`simplify`	should the formula be expanded and simplified, the pre-1.7.0 behaviour?
`...`	further arguments passed to or from other methods.
`allowDotAsName`	normally `.` in a formula refers to the remaining variables contained in `data`. Exceptionally, `.` can be treated as a name for non-standard uses of formulae.

Details

This function is adapted from `terms.formula` in `stats` package and is designed for the use of `bayesglm` and `bayesglm.hierachical`. Unlike `terms.formula`, `terms.bayes` keeps all terms unordered.

Author(s)

Yu-Sung Su ⟨ys463@columbia.edu⟩

See Also

`terms`, `terms.object`, `model.matrix.bayes`, `bayesglm`.

`triangleplot` *Triangle Plot*

Description

Function for making a triangle plot from a square matrix

Usage

```
triangleplot (x, y=NULL, cutpts=NULL, details=TRUE,
              n.col.legend=5, cex.col=0.7,
              cex.var=0.9, digits=1, color=FALSE)
```

Arguments

`x`	a square matrix.
`y`	a vector of names that corresponds to each element of the square matrix x.
`cutpts`	a vector of cutting points for color legend, default is NULL. The function will decide the cutting points if cutpts is not assigned.
`details`	show more than one digits correlaton values. Default is TRUE. FALSE is suggested to get readable output.
`n.col.legend`	number of legend for the color thermometer
`cex.col`	font size of the color thermometer.
`cex.var`	font size of the variable names.
`digits`	number of digits shown in the text of the color theromoeter.
`color`	color of the plot, default is FALSE, which uses gray scale.

Details

The function makes a triangle plot from a square matrix, e.g., the correlation plot, see `corrplot`. If a square matrix contains missing values, the cells of missing values will be marked “x".

Author(s)

Yu-Sung Su ⟨ys463@columbia.edu⟩

See Also

`corrplot`

Examples

```
# create a square matrix
x <- matrix(runif(1600, 0, 1), 40, 40)

# fig 1
triangleplot(x)

# fig 2 assign cutting points
triangleplot(x, cutpts=c(0,0.25,0.5,0.75,1), digits=2)
```

```
# fig 3 if x contains missing value
x[12,13] <- x[13,12] <- NA
x[25,27] <- x[27,25] <- NA
triangleplot(x)
```

Index