

# Package ‘arm ’

January 9, 2007

**Version** 1.0-4

**Date** 2007-1-9

**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, BRugs

**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 . . . . .	2
bayesglm . . . . .	3
bayespolr . . . . .	6
coefplot . . . . .	9
corrplot . . . . .	11
display . . . . .	13
go . . . . .	15
invlogit . . . . .	16
lalonge . . . . .	17
matching . . . . .	18
mcsamp . . . . .	19
rescale . . . . .	21
se.coef . . . . .	22
sigma.hat . . . . .	23
sim . . . . .	24
standardize . . . . .	26

<b>Index</b>	<b>28</b>
--------------	-----------

balanceplot

*Plot of Balance Statistics***Description**

This function plots the balance statistics before and after matching.

**Usage**

```
balanceplot(rawdata, matched, pscore.fit,
            longcovnames = NULL,
            main = "Standardized Difference in Means",
            cex.main = 1, cex.vars = 0.8, cex.pts = 0.8,
            mar = c(0, 15, 4, 2), mgp = c(2, 0.25, 0),
            oma = c(0, 0, 0, 0), tcl = -0.2, ...)
```

**Arguments**

rawdata	unmatched rawdata
matched	matched data using "matching" function
pscore.fit	glm.fit object to get pscore
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.

**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 + 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(lalonde, 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,
          scaled = 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, 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 <a href="#">family</a> for details of family functions.)

<code>data</code>	an optional data frame, list or environment (or object coercible by <code>as.data.frame</code> to a data frame) containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>glm</code> is called.
<code>weights</code>	an optional vector of weights to be used in the fitting process. Should be <code>NULL</code> or a numeric vector.
<code>subset</code>	an optional vector specifying a subset of observations to be used in the fitting process.
<code>na.action</code>	a function which indicates what should happen when the data contain NAs. The default is set by the <code>na.action</code> setting of <code>options</code> , and is <code>na.fail</code> if that is unset. The “factory-fresh” default is <code>na.omit</code> . Another possible value is <code>NULL</code> , no action. Value <code>na.exclude</code> can be useful.
<code>start</code>	starting values for the parameters in the linear predictor.
<code>etastart</code>	starting values for the linear predictor.
<code>mustart</code>	starting values for the vector of means.
<code>offset</code>	this can be used to specify an <i>a priori</i> known component to be included in the linear predictor during fitting. This should be <code>NULL</code> or a numeric vector of length either one or equal to the number of cases. One or more <code>offset</code> terms can be included in the formula instead or as well, and if both are specified their sum is used. See <code>model.offset</code> .
<code>control</code>	a list of parameters for controlling the fitting process. See the documentation for <code>glm.control</code> for details.
<code>model</code>	a logical value indicating whether <i>model frame</i> should be included as a component of the returned value.
<code>method</code>	the method to be used in fitting the model. The default method <code>"glm.fit"</code> uses iteratively reweighted least squares (IWLS). The only current alternative is <code>"model.frame"</code> which returns the model frame and does no fitting.
<code>x, y</code>	For <code>glm</code> : 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 <code>glm.fit</code> : <code>x</code> is a design matrix of dimension $n \times p$ , and <code>y</code> is a vector of observations of length <code>n</code> .
<code>contrasts</code>	an optional list. See the <code>contrasts.arg</code> of <code>model.matrix.default</code> .
<code>intercept</code>	logical. Should an intercept be included in the <i>null</i> model?
<code>prior.mean</code>	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.
<code>prior.scale</code>	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.
<code>prior.scale.for.intercept</code>	prior scale for the intercept: default is 10.
<code>prior.df</code>	for t distribution: default is 1 (Cauchy). Set to <code>Inf</code> 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.
<code>scaled</code>	if <code>scaled = TRUE</code> , then the prior distribution is rescaled: default is <code>TRUE</code>
<code>n.iter</code>	default is 100.
<code>...</code>	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., `offsets`, `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⟩](mailto:gelman@stat.columbia.edu); Yu-Sung Su [⟨ys463@columbia.edu⟩](mailto:ys463@columbia.edu); Maria Grazia Pittau [⟨grazia@stat.columbia.edu⟩](mailto:grazia@stat.columbia.edu); Aleks Jakulin [⟨Jakulin@stat.columbia.edu⟩](mailto: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) # (using the display() function from regression.R)

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)

```

---

 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"),

```

```
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

<code>formula</code>	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.
<code>data</code>	an optional data frame in which to interpret the variables occurring in 'formula'.
<code>weights</code>	optional case weights in fitting. Default to 1.
<code>start</code>	initial values for the parameters. This is in the format 'c(coefficients, zeta)'
<code>...</code>	additional arguments to be passed to 'optim', most often a 'control' argument.
<code>subset</code>	expression saying which subset of the rows of the data should be used in the fit. All observations are included by default.
<code>na.action</code>	a function to filter missing data.
<code>contrasts</code>	a list of contrasts to be used for some or all of the factors appearing as variables in the model formula.
<code>Hess</code>	logical for whether the Hessian (the observed information matrix) should be returned.
<code>model</code>	logical for whether the model matrix should be returned.
<code>method</code>	logistic or probit or complementary log-log or cauchit (corresponding to a Cauchy latent variable and only available in R >= 2.1.0).
<code>prior.mean</code>	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.
<code>prior.scale</code>	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.
<code>prior.df</code>	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.
<code>scaled</code>	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.
<code>prior.mean.for.cutpoints</code>	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.
<code>prior.scale.for.cutpoints</code>	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.

<code>prior.df.for.cutpoints</code>	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.
<code>n.iter</code>	default is 100.

### Details

The program is a simple alteration of `polr()` 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⟩](mailto:gelman@stat.columbia.edu); Yu-Sung Su [⟨ys463@columbia.edu⟩](mailto:ys463@columbia.edu); Maria Grazia Pittau [⟨grazia@stat.columbia.edu⟩](mailto: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)
```



coefplot

*Function for Making Coefficient Plot***Description**

A function that plot the coefficients from a lm or glm fit

**Usage**

```
coefplot(fit, longnames=NULL,
         xlim=NULL, ylim=NULL,
         xlab="", ylab="", main="", ...,
         intercept=FALSE, varnames=TRUE, display.fit=FALSE,
         cex.var=0.8, cex.pts=0.9)
```

**Arguments**

fit	fitted objects-lm or glm.
longnames	long variable names (vector). If not provided, use default variable names.
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'.
...	other plot options may be passed to this function, see <a href="#">plot.default</a> and <a href="#">par</a> .
intercept	If TRUE will plot intercept, default=FALSE.
varnames	If FALSE will not plot variable names, default=TRUE.
display.fit	If TRUE will display fitted output, default=FALSE.
cex.var	The fontsize of the variable names, default=0.8.
cex.pts	The size of data points, default=0.9.

**Details**

This function plots coefficients from lm and glm 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)

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

# plot 3
par (mar=c(2,2,2,2))
coefplot(fit2, longnames=longnames, varnames=FALSE)

# plot 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(0,7.5), 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)
```

corrplot

*Function for Making Correlation Plot***Description**

Function for making a correlation plot starting from a data matrix

**Usage**

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

**Arguments**

data	a data matrix
var.names	variable names of the data matrix, if not provided use default variable names
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
abs	if TRUE, transform all correlation values into positive values.
...	other plot options may be passed to this function, see <a href="#">plot.default</a> and <a href="#">par</a> .

**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, NO. 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, NO. 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,n.col.legend=7)

# example 3
data(lalonde)
corrplot(lalonde, details=FALSE, color=TRUE)
```

**Description**

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

**Usage**

```
display (object, digits = 2)

display.lm (object, digits=2)
display.glm (object, digits=2)
display.mer (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.
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,  $y = a + bx + \text{error}$ ,
# 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)

```

go

*Function to Recall Last Source File***Description**

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

**Usage**

```
go(..., which.source=NULL, add=FALSE, timer=FALSE)
```

**Arguments**

<code>...</code>	list of filenames as character strings
<code>which.source</code>	the source file of which the first argument belongs to
<code>add</code>	add these names to the current list? if replace, then FALSE
<code>timer</code>	time the execution time of <code>go()</code>

**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

```

---

`invlogit`*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:  $\text{logit}^{-1}(x) = e^x / (1 + e^x)$  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⟩](mailto:gelman@stat.columbia.edu), M.Grazia Pittau [⟨grazia@stat.columbia.edu⟩](mailto: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)
```



---

lalonge*Lalonge Dataset*

---

### Description

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

### Usage

```
data(lalonge)
```

### 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 marital 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/lalonge\\_exper\\_04feb2.m](http://elsa.berkeley.edu/~imbens/matlab/lalonge_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	<i>Matching</i>
----------	-----------------

---

**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 + 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(lalonde, matched, fit)
```

mcsamp

*Quick Function to Run mcmcsamp() in lme4***Description**

The 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 generic generates a sample from the posterior distribution of the parameters of a fitted model using Markov Chain Monte Carlo methods.

**Value**

An object of (S3) class "mcmc" suitable for use with the functions in the "coda" 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.

## See Also

[display](#), [sim](#)

## Examples

```
# Here's a simple example of a model of the form,  $y = a + bx + \text{error}$ ,
# 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)
```

---

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, "full12")
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)

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 coef and se.ranef gives a list of standard errors for coef.

## 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,  $y = a + bx + \text{error}$ ,
# 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)
```

**Arguments**

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

**Author(s)**

Andrew Gelman [⟨gelman@stat.columbia.edu⟩](mailto:gelman@stat.columbia.edu); Yu-Sung Su [⟨ys463@columbia.edu⟩](mailto: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 <- glm (y2 ~ x, family=binomial(link="logit"))
sigma.hat (M2)

M3 <- lmer (y1 ~ (1+x|group))
sigma.hat (M3)

M4 <- lmer (y2 ~ (1+x|group), family=binomial(link="logit"))
sigma.hat (M4)
```

**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, n.sims = 100)

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

**Arguments**

<code>object</code>	the output of a call to "lm" with n data points and k predictors
<code>n.sims</code>	number of independent simulation draws to create

**Value**

<code>sigma.sim</code>	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)
<code>beta.sim</code>	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
# Set up the fake data for Example 3
y <- rbinom (n, 1, invlogit (a[group] + b*x))
E3 <- lmer (y ~ x + (1 | group), family=binomial(link="logit"))
display(E3)
E3.sim <- sim (E3)

```

---

standardize	<i>Function for Standardizing Regression Predictors by Centering and Dividing by 2 sd's</i>
-------------	---

---

## 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)
```

# Index

## \*Topic **datasets**

lalonge, 16

## \*Topic **dplot**

balanceplot, 1

coefplot, 8

corrplot, 11

## \*Topic **methods**

balanceplot, 1

bayesglm, 3

bayespolr, 6

display, 12

go, 15

matching, 17

mcsamp, 18

se.coef, 21

sigma.hat, 23

sim, 24

standardize, 25

## \*Topic **models**

bayesglm, 3

bayespolr, 6

invlogit, 15

matching, 17

mcsamp, 18

rescale, 20

se.coef, 21

sim, 24

standardize, 25

as.data.frame, 3

balanceplot, 1, 17, 18

bayesglm, 3, 8, 9

bayespolr, 5, 6

coef, 22

coefplot, 8

cor, 12

corrplot, 11

display, 9, 12, 19, 22–24

display, glm-method(*display*), 12

display, lm-method(*display*), 12

display, mer-method(*display*), 12

display, polr-method(*display*), 12

display.glm(*display*), 12

display.lm(*display*), 12

display.mer(*display*), 12

display.polr(*display*), 12

family, 3

GenMatch, 17

glm, 5, 9, 13, 23, 24

glm.control, 4

go, 15

invlogit, 15

lalonge, 16

lm, 9, 13, 23, 24

lmer, 13, 23, 24

matching, 2, 17, 17

mcsamp, 18, 24

model.offset, 4

na.exclude, 3

na.fail, 3

na.omit, 3

offset, 4

options, 3

par, 2, 9, 11, 12

plot.default, 9, 11

polr, 8, 13

rescale, 20, 26

se.coef, 21

se.coef, glm-method(*se.coef*), 21

se.coef, lm-method(*se.coef*), 21

se.coef, mer-method(*se.coef*), 21

se.coef.glm(*se.coef*), 21

se.coef.lm(*se.coef*), 21

se.coef.mer(*se.coef*), 21

se.fixef(*se.coef*), 21

se.ranef(*se.coef*), 21

`sigma.hat`, 22, 23  
`sigma.hat`, `glm-method(sigma.hat)`,  
23  
`sigma.hat`, `lm-method(sigma.hat)`,  
23  
`sigma.hat`, `mer-method(sigma.hat)`,  
23  
`sigma.hat.glm(sigma.hat)`, 23  
`sigma.hat.lm(sigma.hat)`, 23  
`sigma.hat.mer(sigma.hat)`, 23  
`sim`, 19, 24  
`sim`, `glm-method(sim)`, 24  
`sim`, `lm-method(sim)`, 24  
`sim`, `mer-method(sim)`, 24  
`sim.glm(sim)`, 24  
`sim.lm(sim)`, 24  
`sim.mer(sim)`, 24  
`standardize`, 21, 25  
`summary`, 13, 23