

stats (version 3.6.2)

model.matrix: Construct Design Matrices

Description

``model.matrix`` creates a design (or model) matrix, e.g., by expanding factors to a set of dummy variables (depending on the contrasts) and expanding interactions similarly.

Usage

```
model.matrix(object, ...)  
  
# S3 method for default  
model.matrix(object, data = environment(object),  
              contrasts.arg = NULL, xlev = NULL, ...)
```

Arguments

object	an object of an appropriate class. For the default method, a model formula or a <code>`terms`</code> object.
data	a data frame created with <code>`model.frame`</code> . If another sort of object, <code>`model.frame`</code> is called first.
contrasts.arg	a list, whose entries are values (numeric matrices, <code>`function`</code> s or character strings naming functions) to be used as replacement values for the <code>`contrasts`</code> replacement function and whose names are the names of columns of <code>`data`</code> containing <code>`factor`</code> s.
xlev	to be used as argument of <code>`model.frame`</code> if <code>`data`</code> is such that <code>`model.frame`</code> is called.
...	further arguments passed to or from other methods.

Value

The design matrix for a regression-like model with the specified formula and data.

There is an attribute `"assign"`, an integer vector with an entry for each column in the matrix giving the term in the formula which gave rise to the column. Value `0` corresponds to the intercept (if any), and positive values to terms in the order given by the `term.labels` attribute of the `terms` structure corresponding to `object`.

If there are any factors in terms in the model, there is an attribute `"contrasts"`, a named list with an entry for each factor. This specifies the contrasts that would be used in terms in which the factor is coded by contrasts (in some terms dummy coding may be used), either as a character vector naming a function or as a numeric matrix.

Details

`model.matrix` creates a design matrix from the description given in `terms(object)`, using the data in `data` which must supply variables with the same names as would be created by a call to `model.frame(object)` or, more precisely, by evaluating `attr(terms(object), "variables")`. If `data` is a data frame, there may be other columns and the order of columns is not important. Any character variables are coerced to factors. After coercion, all the variables used on the right-hand side of the formula must be logical, integer, numeric or factor.

If `contrasts.arg` is specified for a factor it overrides the default factor coding for that variable and any `"contrasts"` attribute set by `C` or `contrasts`. Whereas invalid `contrasts.arg`'s have been ignored always, they are warned about since R version 3.6.0.

In an interaction term, the variable whose levels vary fastest is the first one to appear in the formula (and not in the term), so in `~ a + b + b:a` the interaction will have `a` varying fastest.

By convention, if the response variable also appears on the right-hand side of the formula it is dropped (with a warning), although interactions involving the term are retained.

References

Chambers, J. M. (1992) *Data for models*. Chapter 3 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

See Also

`model.frame`, `model.extract`, `terms`

``sparse.model.matrix`` from package `Matrix` for creating *sparse* model matrices, which may be more efficient in large dimensions.

Examples

```
# NOT RUN {
ff <- log(Volume) ~ log(Height) + log(Girth)
utils::str(m <- model.frame(ff, trees))
mat <- model.matrix(ff, m)

dd <- data.frame(a = gl(3,4), b = gl(4,1,12)) # balanced 2-way
options("contrasts") # typically 'treatment' (for unordered factors)
model.matrix(~ a + b, dd)
model.matrix(~ a + b, dd, contrasts = list(a = "contr.sum"))
model.matrix(~ a + b, dd, contrasts = list(a = "contr.sum", b = contr.poly))
m.orth <- model.matrix(~a+b, dd, contrasts = list(a = "contr.helmert"))
crossprod(m.orth) # m.orth is ALMOST orthogonal
# invalid contrasts.. ignored with a warning:
stopifnot(identical(
  model.matrix(~ a + b, dd),
  model.matrix(~ a + b, dd, contrasts.arg = "contr.F00")))
# }
```

Run this code

Run the code above in your browser using [DataLab](#)

Powered by [DataCamp](#)