The repolr Package

June 20, 2008

Type Package

Title Repeated measures proportional odds logistic regression

Version 1.0

Date 2008-05-01

Author Nick Parsons

Maintainer Nick Parsons <nick.parsons@warwick.ac.uk>

Depends gee

Description Package fits regression model to repeated ordinal response variables using generalized estimating equations.

License GPL

URL http://www.warwick.ac.uk/go/repolr

R topics documented:

HHSpain *Harris Hip Pain Scores*

Description

Harris hip pain scores, measured on 4 a point ordinal score scale, post-surgery for 58 patients at 3 occasions.

Usage

```
data(HHSpain)
```

Format

A data frame with 174 observations on the following 4 variables.

`Patient` a patient identifier variable.

`Sex` a factor with levels `F M`

`Time` recorded at baseline (1), two years (2) and five years (5) post-surgery.

`HHSpain` a pain score coded as none (1), slight (2), mild (3) and moderate or marked (4)

Source

Insert reference when paper is published.

Examples

```
data(HHSpain)
str(HHSpain)
```

achilles *Achilles Tendon Rupture*

Description

EuroQol activity scores, measured on 3 a point ordinal score scale, post-surgery for 48 patients at 3 occasions.

Usage

```
data(achilles)
```

Format

A data frame with 144 observations on the following 4 variables.

`Patient` A patient identifier variable.

`Treat` Post-surgery treatments are either immediate mobilisation in a carbon-fibre orthosis with three 1.5cm heel raises (1) or traditional plaster cast immobilisation (2).

`Time` Recorded at baseline (1), six months (2) and one year (3) post-surgery.

`Activity` ability to undertake usual activities post-surgery; this was scored by each patient as either no problem (1), some problem (2) or an inability to perform (3) usual activity (e.g. work, leisure, housework etc).

Source

Costa, M.L., MacMillan, K., Halliday, D., Chester, R., Shepstone, L., Robinson, A.H.N., Donell, S.T. (2006). Randomised controlled trials of immediate weight-bearing mobilisation for rupture of the tendon Achillis. *Journal of Bone and Joint Surgery (British)* 88-B, 69-77.

Examples

```
data(achilles)
str(achilles)
```

---

ogee *Internal Function to fit Proportional Odds Model*

Description

An internal function called by `repolr`, that would not normally be called by user. This is a modified version of `gee` for fitting the proportional odds model.

Usage

```
ogee(formula = formula(data), id = id, data = parent.frame(), subset, na.action, R
```

Arguments

`formula`	as `gee`
`id`	as `gee`
`data`	as `gee`
`subset`	as `gee`
`na.action`	as `gee`
`R`	as `gee`
`b`	as `gee`
`tol`	as `gee`

maxiter	as gee
family	as gee
corstr	as gee
Mv	as gee
silent	as gee
contrasts	as gee
scale.fix	as gee
scale.value	as gee
v4.4compat	as gee

Details

Not generally to be called by user.

`ord.expand` *Expand Ordinal Data for GEE Modeling*

Description

Expands ordinal score data into appropriate form, K-1 new binary variables for ordinal score with K categories, for model fitting using GEE.

Usage

```
ord.expand(scores, data, subjects, categories)
```

Arguments

scores	a character string identifying the response variable.
data	a dataframe in which to interpret the variables occurring in the formula.
subjects	a character string specifying the name of the subject variable.
categories	a numeric variable indicating the number of ordinal score categories.

Details

Mainly for internal use with `repolr`.

Value

An expanded dataframe.

repolr-package *Repeated Measures Proportional Odds Logistic Regression using GEE*

Description

The package allows regression models to be fitted to repeated ordinal scores, for the proportional odds model, using a modified version of the generalized estimating equation (GEE) method. The algorithm used estimates the correlation parameter by minimizing the generalized variance of the regression parameters at each step of the fitting algorithm. Parameter estimation is available for the uniform and first-order autoregressive correlation models, for data potentially recorded at irregularly spaced time intervals and with arbitrary patterns of missingness. A sensitivity analysis is used to assess the importance of the correlation model and a score test is available that tests the assumption of proportional odds.

Usage

```
repolr(formula, subjects, data, times, categories, scalevalue = 1, corstr = "ar1",
```

Arguments

`formula`	a formula, as for other regression models.
`subjects`	a character string specifying the name of the subject variable.
`data`	a dataframe in which to interpret the variables occurring in the formula.
`times`	a vector of times which occur within subject clusters; e.g. for four evenly spaced times c(1,2,3,4).
`categories`	a numeric variable indicating the number of ordinal score categories.
`scalevalue`	a numeric variable giving the value to which the scale parameter should be fixed.
`corstr`	a character string specifying the correlation structure. The following are permitted: "`ar1`", "`uniform`" and "`independence`".
`maxiter`	the maximum number of iterations.
`tol`	the tolerance used in the fitting algorithm.
`alpha`	an initial value for the correlation parameter.
`po.test`	a logical variable; if true a score test for proportional odds is reported.
`fixed`	a logical variable; if true the correlation is fixed at the initial value (`alpha`) during model fitting.

Details

The `repolr` function fits generalized estimating equations using GEE and `srepolr` can be used to assess the significance of the estimated correlation parameter.

The user is required to specify: (i) a data set name (`data`), (ii) a model formula (`formula`), (iii) a cluster identification variable (`subjects`), (iv) a time variable (`time`) and (v) the number of categories used for the response variable (`categories`). The data set may contain records with

missing data for either the response variable or the explanatory variables. The response variable must have at least three ordered categories (K greater than or equal to 3) and, as K cut-point parameters are estimated, an additional intercept term can not be explicitly included in the model formula. A subject variable, which takes integer values from 1 to N with no missing values allowed, indicates the data clusters (patients or experimental units) and a time variable indicates the within cluster ordering; times must be ordered integers starting from one and spaced to indicate the relative distance between successive times. For instance, four observations at equally spaced times would be entered as 1, 2, 3 and 4, whereas if the first two observations were separated by half the time interval of the other observations then coding would be 1, 2, 4 and 6. The data must be pre-sorted by time clusters within each subject, and complete, i.e. where data is missing for a particular time it must be indicated as such. The available options for the correlation model (`corstr`) are AR1, uniform, fixed and independence, with default setting AR1.

Additionally there are a number of other algorithm related options. A fixed scale parameter (`scalevalue`) is assumed to be 1, but otherwise it can be set to any positive value. The algorithm is generally robust to the initial value for alpha (default setting = 0.5), however a starting value for alpha can be set. The maximum number of iterations allowed (`maxiter`; default setting = 10) can be specified and convergence (`tol`) is monitored using the average relative change in the regression parameters at each iteration of the algorithm; the default setting (0.001) is such that the algorithm continues until this drops to below 0.1 percent. The algorithm will fail if the working correlation matrix is not positive definite and, in order to speed convergence for the AR1 and uniform correlation models, is constrained to be between 0.05 and 0.95. The proportional odds assumption can be tested, using a score test, by setting the option `po.test` to TRUE (default).

The functions require the user to have pre-installed the latest version of the R GEE solver `gee` and the fitted model `fitted.mod[[gee]]` inherits the class attributes of this package; correlation model ouput is available from `fitted.mod[[corr]]`.

Value

The fitted model is an object of class "`gee`" and the correlation model has following values in addition to the model set-up variables:

`crit`	convergence criterion.
`score.test`	score test statistic.
`iter`	number of iterations.
`alpha`	estimate of the correlation parameter.
`grad1`	first derivative of generalized variance at convergence.
`grad2`	second derivative of generalized variance at convergence.

References

Parsons, N.R., Costa, M.L., Achten, J. and Stallard, N. (2007). Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R. (Under Revision; *Computational Statistics and Data Analysis*)

Parsons, N.R., Edmondson, R.N. and Gilmour, S.G. (2006). A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research.

Applied Statistics, 55, 507-524.

Stiger, T.R., Barnhart, H.X., Williamson, J.M. (1999). Testing proportionality in the proportional odds model fitted with GEE. *Statistics in Medicine* 18, 1419-1433.

See Also

srepolr

Examples

```
## fit uniform correlation model to achilles tendon rupture data
data(achilles)
fitted.mod <- repolr(formula=Activity~factor(Treat)*Time, subjects="Patient",
data=achilles, categories=3, times=c(1,2,3), corstr="uniform", tol=0.001,
scalevalue=1, alpha=0.5, po.test=TRUE, fixed=FALSE)
summary(fitted.mod[["gee"]])
fitted.mod[["corr"]]

## fit ar1 correlation model to hip resurfacing pain data
data(HHSpain)
fitted.mod <- repolr(HHSpain~factor(Sex)*factor(Time), subjects="Patient",
data=HHSpain, categories=4, times=c(1,2,5), corstr="ar1", alpha=0.5)
summary(fitted.mod[["gee"]])

## plot sensitivity
sensit <- vector(length=81); alphas <- seq(0.1,0.9,0.01)
for (j in 1:81){
sensit[j] <- srepolr(mod.gee=fitted.mod,alpha=alphas[j],data=HHSpain)
}
Nsensit <-sensit/
srepolr(mod.gee=fitted.mod, alpha=fitted.mod[["corr"]]$alpha, data=HHSpain)
plot(x=alphas,y=Nsensit,type="l",lwd=2)
```

`srepolr` *Explore Sensitivity of repolr*

Description

Determines the logarithm of the determinant of the covariance matrix of the regression parameters (the logarithm of the generalized variance) from a previous fit of `repolr` for an arbitrary correlation parameter.

Usage

```
srepolr(data, mod.gee, alpha = 0.5)
```

Arguments

data	a dataframe in which to interpret the model fitted by repolr
mod.gee	a fitted model from repolr
alpha	a value for the correlation parameter

Value

Returns logarithm of the determinant of the covariance matrix.

Examples

```
## fit ar1 correlation model to hip resurfacing pain data
data(HHSpain)
fitted.mod <- repolr(HHSpain~factor(Sex)*factor(Time), subjects="Patient",
data=HHSpain, categories=4, times=c(1,2,5), corstr="ar1", alpha=0.5)

## plot sensitivity
sensit <- vector(length=81); alphas <- seq(0.1,0.9,0.01)
for (j in 1:81){
sensit[j] <- srepolr(mod.gee=fitted.mod,alpha=alphas[j],data=HHSpain)
}
Nsensit <-sensit/
srepolr(mod.gee=fitted.mod, alpha=fitted.mod[["corr"]]$alpha, data=HHSpain)
plot(x=alphas,y=Nsensit,type="l",lwd=2)
```

Index