Package ‘NominalLogisticBiplot’

February 19, 2015

Type Package

Title Biplot representations of categorical data

Version 0.2

Date 2014-05-01

Author Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

Description Analysis of a matrix of polytomous items using Nominal Logistic Biplots (NLB)
according to Hernandez-Sanchez and Vicente-Villardon (2013).
The NLB procedure extends the binary logistic biplot to nominal (polytomous) data.
The individuals are represented as points on a plane and the variables are represented
as convex prediction regions rather than vectors as in a classical or binary biplot.
Using the methods from Computational Geometry, the set of prediction regions is converted to a set of points
in such a way that the prediction for each individual is established by its closest
``category point''. Then interpretation is based on distances rather than on projections.
In this package we implement the geometry of such a representation and construct computational algorithms
for the estimation of parameters and the calculation of prediction regions.

License GPL (>= 2)

Encoding latin1

Repository CRAN

Depends R (>= 2.15.1),mirt,gmodels,MASS

LazyData yes

Archs i386, x64

NeedsCompilation no

Date/Publication 2014-05-02 07:13:20

R topics documented:

`NominalLogisticBiplot-package`

Nominal Logistic Biplot representations for polytomous data

Description

Analysis of a matrix of polytomous items using Nominal Logistic Biplots (NLB) according to Hernandez-Sanchez & Vicente-Villardon (2013). The NLB procedure extends the binary logistic biplot to nominal (polytomous) data. The individuals are represented as points on a plane and the variables are represented as convex prediction regions rather than vectors as in a classical or binaly biplot. Using the methods from the Computational Geometry, the set of prediction regions is converted to a set of points in such a way that the prediction for each individual is established by its closest "category point". Then interpretation is based on distances rather than on projections. In this package we implement the geometry of such a representation and construct computational algorithms for the estimation of parameters and the calculation of prediction regions

Details

Package:	NominalLogisticBiplot
Type:	Package
Version:	1.0
Date:	2013-08-05
License:	GPL (>=2)

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

See Also

NominalLogisticBiplot,NominalLogBiplotEM,multiquad,summary.nominal.logistic.biplot,plot.nominal.logis

Examples

```
data(HairColor)
nlbo = NominalLogisticBiplot(HairColor,sFormula=NULL,numFactors=2,
method="EM",penalization=0.2,show=FALSE)
summary(nlbo)
plot(nlbo,QuitNotPredicted=TRUE,ReestimateInFocusPlane=TRUE,
planex = 1,planey = 2,proofMode=TRUE,LabelInd=TRUE,AtLeastR2 = 0.01
,xlimi=-1.5,xlimu=1.5,ylimi=-1.5,ylimu=1.5,linesVoronoi = TRUE
,SmartLabels = FALSE, PlotInd=TRUE,CexInd = c(0.6,0.7,0.5,0.4,0.5,0.6,0.7)
,PchInd = c(1,2,3,4,5,6,7),ColorInd="black",PlotVars=TRUE,LabelVar = TRUE
,PchVar = c(1,2,3,4,5),ColorVar = c("red","black","yellow","blue","green")
,ShowResults=TRUE)
```

---

afc *Simple Correspondence Analysis*

Description

This function calculates simple correspondence analysis for a data matrix.

Usage

```
afc(x, dim = 2, alpha = 1)
```

Arguments

x	A frequency matrix or a binary matrix obtained from the original data set of nominal variables.
dim	Number of dimensions for the solution
alpha	Biplot weight for rows and columns. 1 means rows in principal coordinates and columns in standard coordinates, 0 means rows in standard coordinates and columns in principal coordinates.

Value

An object of class "afc.sol".This has some components:

- `Title` Title of the statistical technique
- `Non_Scaled_Data` Original data
- `Minima` vector with the minimum values for each column of the initial data matrix
- `Maxima` vector with the maximum values for each column of the initial data matrix
- `Initial_Transformation` Name of the transformation for the data
- `Scaled_Data` Scaled data according to the transformation
- `nrows` Number of rows of the data set
- `ncols` Number of columns of the data set
- `dim` Number of dimensions for the solution
- `CumInertia` Acumulated Inertia
- `Scale_Factor` Scale factor for the transformation
- `RowCoordinates` Coordinates for the individuals in the reduced dimension space
- `ColCoordinates` Coordinates for the variables in the reduced dimension space
- `RowContributions` Contributions of the dimensions to explain the inertia of each row
- `ColContributions` Contributions of the dimensions to explain the inertia of each column
- `Inertia` Inertia for each dimension
- `Eigenvalues` Eigenvalues

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

References

BENZECRI, J.P. (1973) *L'analyse des Donnees*. Vol. 2. L'analyse des correspondences. Dunod. Paris.

See Also

NominalMatrix2Binary

Examples

```
  data(HairColor)
  G = NominalMatrix2Binary(data.matrix(HairColor))
mca=afc(G,dim=2)
  mca
```

Env *Ecological Factors in Farm Management.*

Description

The farms Env data frame has 20 rows and 4 columns. The rows are farms on the Dutch island of Terschelling and the columns are factors describing the management of grassland.

Usage

```
data(Env)
```

Format

This data frame contains the following columns:

Mois five levels of soil moisture,although level 3 does not occur in the data. Levels are labelled M1, M2, M4 and M5.

Manag Grassland management type (SF = standard farming, BF = biological farming, HF = hobby farming, NM = nature conservation management).

Use Grassland use (U1 = it exists production, U2 = intermediate,U3 = grazing).

Manure Manure usage (C0,C1,C2,C3 and C4)

Source

J.C. Gower and D.J. Hand (1996) *Biplots*. Chapman & Hall, Table 4.6.

Quoted as from:
R.H.G. Jongman, C.J.F. ter Braak and O.F.R. van Tongeren (1987) *Data Analysis in Community and Landscape Ecology.* PUDOC, Wageningen.

References

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S.* Fourth edition. Springer.

Examples

```
data(Env)
```

`Generators` *Generators (points) of the tesselation generated by a nominal variable.*

Description

With the parameters resulting from fitting a nominal logistic model to the row scores for a given variable, the function calculates all the information necessary to plot the tessellation generated by the fit. The final user will not normally use this function.

Usage

```
Generators(beta)
```

Arguments

`beta`	Matrix with the estimated parameters for a given nominal variable. It has as many rows as the number of categories minus one and three columns (one for the constant and other two for the x-y coordinates on the plane).

Value

An object of class `"voronoiprob"`. This has the components:

`x`	x-coordinates for the real points (Vertices of the tessellation).
`y`	y-coordinates for the real points (Vertices of the tessellation).
`n1`	vector with the first neighbours of the real points
`n2`	vector with the second neighbours of the real points
`n3`	vector with the third neighbours of the real points
`dummy.x`	x-coordinates for the dummy points
`dummy.y`	y-coordinates for the dummy points
`ndummy`	Number of dummies
`IndReal`	Matrix with the indices of each real point in the tessellation
`Centers`	Matrix with the points resulting from inverting the tessellation
`hideCat`	Vector to indicate if there are some hidden categories
`equivRegiones`	Matrix with the new re-numbered categories (when some are hidden)

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

References

Hern\'andez S\'anchez, J. C., & Vicente-Villard\'on, J. L. (2013). Logistic biplot for nominal data. arXiv preprint arXiv:1309.5486.

Gower, J. & Hand, D. (1996), *Biplots, Monographs on statistics and applied probability* 54. London: Chapman and Hall., 277 pp.

Evans, D. & Jones, S. (1987), *Detecting voronoi (area of influence) polygons* ,Mathematical Geology 19(6), 523–537.

Hartvigsen, D. (1992), *Recognizing voronoi diagrams with linear programming*, ORSA Journal on Computing 4, 369–374.

Schoenberg, F., Ferguson, T. & Li, C. (2003), *Inverting dirichlet tesselations*, Computer journal 46(1), 76–83.

Examples

```
data(HairColor)
data = data.matrix(HairColor)
xEM = NominalLogBiplotEM(data, dim = 2,showResults = FALSE)
nomreg = polylogist(data[,2],xEM$RowCoordinates[,1:2],penalization=0.1)
tesselation = Generators(nomreg$beta)
tesselation
```

HairColor *Demographic Data*

Description

The sample data corresponds to 7 people and shows some demographic characteristics.

Usage

```
data(HairColor)
```

Format

This data frame contains 7 observation for the following 5 columns:

Sex two levels (M=male,F=female)

HairColor four levels of hair color (Dark, Grey, Fair and Brown)

Region (E = England,S = Scotland, W = Wales)

Work (Manual,Clerical,Professional)

Education (School,Univ,Postgrad)

Source

Gower, J., Gardner-Lubbe,S., Le Roux,N. (2011). “Understanding Biplots.” *Wiley*.

Examples

```
data(HairColor)
```

---

hermquad *Gauss-Hermite quadrature*

Description

Computes the Hermite Quadrature weights for a set of grid points

Usage

```
hermquad(N)
```

Arguments

N	Number of nodes for the quadrature

Value

An object of class "GaussQuadrature". This has the components:

X	Coordinates of the nodes
W	Weights asociated to each node

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

References

Stroud, A.H. and Secrest, D. (1966) *Gaussian Quadrature Formulas*, Englewood Cliffs, NJ: Prentice-Hall.

Hildebrand,F. B. (1987) *Intoduction to Numerical Analysis 2nd Ed*, Dover Publications, New York, page 385

Examples

```
hermquad(10)
```

multiquad *Multidimensional Gauss-Hermite quadrature*

Description

This function computes the gauss-hermite quadrature in more than one dimension.

Usage

```
multiquad(nnodos, dims)
```

Arguments

nnodos	Number of nodes.
dims	Number of dimensions of the quadrature

Value

An object of class "MultiGaussQuadrature". This has the components:

X	Coordinates of the nodes
A	Weights asociated to each node

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

References

Jackel, P. (2005) *A note on multivariate Gauss-Hermite quadrature*

http://www.pjaeckel.webspace.virginmedia.com/ANoteOnMultivariateGaussHermiteQuadrature.pdf

See Also

hermquad

Examples

```
multiquad(10,2)
```

`Nominal2Binary` *Transformation of a nominal variable into a binary indicator matrix*

Description

This function transforms a nominal variable into a binary matrix with as many colums as categories. Each row of the matrix has a value of 1 for the corresponding level of the category and 0 elsewhere.

Usage

```
Nominal2Binary(y)
```

Arguments

`y`	A vector containing the values of nominal variable measured on a set of individuals- The values must be integers starting at 1.

Value

An object of type matrix:

`Z`	The binary indicator matrix asociated to the nominal variable

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

Examples

```
data(HairColor)
Nominal2Binary(as.numeric(HairColor[,1]))
```

`NominalDistances` *Distances between indidividuals calculated from nominal variables.*

Description

This function calculates the hamming distances (or similarities) among individuals from a nominal data matrix.

Usage

```
NominalDistances(x, similarities = FALSE)
```

Arguments

`x`	This parameter is a matrix with the nominal variables
`similarities`	Boolean parameter to specify if the user wants a distances matrix or a similarities matrix. By default this parameter is FALSE, so the function calculates the distances.

Value

The function returns a matrix with the distances or similarities

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

References

Boriah, S., Chandola, V. & Kumar,V.(2008) *Similarity measures for categorical data: A comparative evaluation*. In proceedings of the eight SIAM International Conference on Data Mining, pp 243–254

Examples

```
data(HairColor)
NominalDistances(data.matrix(HairColor))
```

`NominalLogBiplotEM` *Alternated EM algorithm for Nominal Logistic Biplots*

Description

This function computes, with an alternated algorithm, the row and column parameters of a Nominal Logistic Biplot for polytomous data. The row coordinates (E-step) are computed using multidimensional Gauss-Hermite quadratures and Expected *a posteriori* (EAP) scores and parameters for each variable or items (M-step)using Ridge Nominal Logistic Regression to solve the separation problem present when the points for different categories of a variable are completely separataed on the representation plane and the usual fitting methods do not converge. The separation problem is present in almost avery data set for which the goodness of fit is high.

Usage

```
NominalLogBiplotEM(x, dim = 2, nnodos = 10, tol = 1e-04,
maxiter = 100, penalization = 0.2,initial=1,alfa=1, Plot = FALSE,
showResults = FALSE)
```

Arguments

x	Matrix with the nominal data. The matrix must be in numerical form.
dim	Dimension of the solution
nnodos	Number of nodes for the multidimensional Gauss-Hermite quadrature
tol	Value to stop the process of iterations.
maxiter	Maximum number of iterations in the process of solving the regression coefficients.
penalization	Penalization used in the diagonal matrix to avoid singularities.
initial	Value to decide the method(1-Correspondence analysis, 2-Mirt) that calculates the initial abilities values for the individuals.
alfa	If initial parameter method is correspondence analysis, this parameter determines the weight for rows and columns.
Plot	Boolean parameter to plot the row coordinates
showResults	Boolean parameter to show all the information about the iterations.

Value

An object of class "nominal.logistic.biplot.EM".This has components:

RowCoordinates
: Coordinates for the individuals in the reduced space

ColumnModels
: List with information about the Nominal Logistic Models calculated for each variable including: estimated parameters with covariances and standard errors, log-likelihood, deviances, percents of correct classifications, pvalues and pseudo-Rsquared measures

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

References

Bock,R. & Aitkin,M. (1981),*Marginal maximum likelihood estimation of item parameters: Aplication of an EM algorithm*, Phychometrika 46(4), 443-459.

Gabriel, K. R. (1998). Generalised bilinear regression. Biometrika, 85(3), 689-700.

Vicente-VillardÃ³n, J. L., Galindo VillardÃ³n, M. P., & BlÃ¡zquez Zaballos, A. (2006). Logistic biplots. Multiple correspondence analysis and related methods. London: Chapman & Hall, 503-521.

Gabriel, K. R., & Zamir, S. (1979). Lower rank approximation of matrices by least squares with any choice of weights. Technometrics, 21(4), 489-498.

See Also

polylogist, multiquad

Examples

```
data(HairColor)
data = data.matrix(HairColor)
xEM = NominalLogBiplotEM(data, dim = 2,showResults = FALSE)
xEM
```

NominalLogisticBiplot *Nominal Logistic Biplot for polytomous data*

Description

Function that calculates the parameters of the Nominal Logistic Biplot according to Hernandez-Sanchez & Vicente-Villardon (2013).

Usage

```
NominalLogisticBiplot(datanom, sFormula = NULL, numFactors = 2,
method = "EM", rotation = "varimax", metfsco = "EAP",
nnodos = 10, tol = 1e-04, maxiter = 100, penalization = 0.1,
cte = TRUE,initial=1,alfa=1, show = FALSE)
```

Arguments

datanom	The data set, it can be a *matrix* with integers or a *data frame* with factors. All variables have to be nominal.
sFormula	This parameter follows the unifying interface for selecting variables from a data frame for a plot, test or model. The most common formula es of type y ~ x1+x2+x3. It has a default value of NULL if not specified.
numFactors	Number of dimensions of the solution. It should be lower than the number of variables. It has a default value of 2.
method	This parameter can be: "EM", "ACM", "MIRT" or "PCOA". Method to compute the row coordinates.
rotation	Rotation method to used with "MIRT" option in "coordinates". No effect fror other options.
metfsco	Calculation method for the fscores with "MIRT" option in "coordinates". No effect fror other options.
nnodos	Number of nodes for gauss quadrature in the EM algorithm.
tol	Tolerance for the EM algorithm.
maxiter	Maximum number of iterations in the EM algorithm.
penalization	Penalization for the ridge regression for each variable.
cte	Include constant in the logistic regression model. Default is TRUE.
initial	Value to decide the method(1-Correspondence analysis, 2-Mirt) that calculates the initial abilities values for the individuals.
alfa	If initial parameter method is correspondence analysis, this parameter determines the weight for rows and columns.
show	Show intermediate copmputations. Default is TRUE.

Details

The general algorithm used is essentially an alternated procedure in which parameters for rows and columns are computed in alternated steps repeated until convergence. Parameters for the rows are calculated by expectation (E-step) or by a external procedure (Multiple Correspondence Analysis or Principal Coordinates Analysis) and parameters for the columns are computed by maximization (M-step), i. e., by Nominal Logistic Regression. When the procedure for Row scores is external, only one iteration is performed and the procedure is called "External Nominal Logistic Biplot". Because the aim of the biplot is the representation

There are several options for the computation:

1.- Using the package **mirt** to obtain the row scores, i. e. using a solution obtained from a latent trait model. The column (item) parameters should be directly used by our biplot procedure but, because of the characteristics of the package that performs a default rotation after parameter estimation, we have to reestimate the item parametes to be coherent to the scores.

2.- Using our implementation of the EM algorithm alternating expected a porteriori scores and Ridge Nominal Logistic Regression for each variable.

3.- Using external coordinates for the rows taken from Multiple Correspondence Analysis or Principal Coordinates Analysis and fitting the response surfaces in just one step.

Equations that define a set of probability response surfaces (one for each category and each variable) are no longer sigmoid as in the binary case (Vicente-Villardon et al. (2006)). This means that the level curves are no longer straight lines and then, prediction of probabilities is not made by projection as in the usual linear biplots. For each variable, define a set of convex polygons that can be interpreted as "prediction" regions in the same way as in Gower & Hand (1996). Each pair of response surfaces defined by intersect in a straight line that, projected onto the space of predictors, is the set of points in which the probability of both categories is the same. Those lines are the candidates to be the edges of the convex polygons defining the prediction regions.

Value

An object of class `"nominal.logistic.biplot"`. This has some components:

`dataSet`	Data set of study with all the information about the name of the levels and names of the variables and individuals
`RowCoords`	Coordinates for the individuals in the reduced space
`VariableModels`	Information of the regression resuls for each variable.
`NumFactors`	Number of dimensions selected for the study
`Method`	Method for calculating the row positions
`Rotation`	Type of rotation if we have chosen mirt coordinates
`Methodfscores`	Method of calculation of the fscores in mirt process
`NumNodos`	Number of nodes for the gauss quadrature in EM algorith
`tol`	Cut point to stop the EM-algorith
`maxiter`	Maximum number of iterations in the EM-algorith
`penalization`	Value for the correction of the ridge regression

`cte`	Boolean value to choose if the model for each variable will have independent term
`show`	Boolean value to indicate if we want to see the results of our analysis

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

References

Hernandez, J. C. & Vicente-Villardon, J. L. (2013) Logistic Biplots for Nominal Data. Submitted. Preprint available at : https://www.researchgate.net/publication/256428288_Logistic_Biplot_for_Nominal_Data?ev=prf_pub

Vicente-Villardon, J., Galindo, M.P & Blazquez-Zaballos, A. (2006), *Logistic biplots*,Multiple Correspondence Analysis and related methods pp. 491–509.

Demey, J., Vicente-Villardon, J. L., Galindo, M.P. & Zambrano, A. (2008) *Identifying Molecular Markers Associated With Classification Of Genotypes Using External Logistic Biplots*. Bioinformatics, 24(24), 2832-2838.

Baker, F.B. (1992): Item Response Theory. Parameter Estimation Techniques. Marcel Dekker. New York.

Gabriel, K. (1971), *The biplot graphic display of matrices with application to principal component analysis.*, Biometrika 58(3), 453–467.

Gabriel, K. R. (1998), *Generalised bilinear regression*, Biometrika 85(3), 689–700.

Gabriel, K. R. & Zamir, S. (1979), *Lower rank approximation of matrices by least squares with any choice of weights*, Technometrics 21(4), 489–498.

Gower, J. & Hand, D. (1996), *Biplots, Monographs on statistics and applied probability*. 54. London: Chapman and Hall., 277 pp.

Chalmers,R,P (2012). *mirt: A Multidimensional Item Response Theory Package for the R Environment*. Journal of Statistical Software, 48(6), 1-29. URL http://www.jstatsoft.org/v48/i06/.

See Also

`NominalLogBiplotEM`, `afc`, `PCoA`

Examples

```
data(HairColor)
nlbo = NominalLogisticBiplot(HairColor,sFormula=NULL,
numFactors=2,method="EM",penalization=0.2,show=FALSE)
nlbo

#data(PhD_nomCyL)
#cyl = NominalLogisticBiplot(PhD_nomCyL,sFormula=NULL,
#numFactors=2,method="EM",initial = 1,penalization=0.3,show=FALSE)
#summary(nlboPhD)
#plot(nlboPhD,QuitNotPredicted=TRUE,ReestimateInFocusPlane=TRUE,
```

```
#      planex = 1,planey = 2,proofMode=TRUE,LabelInd=FALSE,AtLeastR2 = 0.01,
#     xlimi=-1.5,xlimu=1.5,ylimi=-1.5,ylimu=1.5,linesVoronoi = TRUE,SmartLabels = FALSE,
#      PlotInd=TRUE,
#      CexInd = c(0.4),
#      PchInd = c(1),
#      ColorInd="azure3",
#      PlotVars=TRUE,LabelVar = TRUE,
#      PchVar = c(1,2,3,4,5,6,7,8,9),ColorVar = c("red","black","maroon","blue","green",
#      "chocolate4","coral3","brown","brown2"),
#      ShowResults=TRUE)
```

NominalMatrix2Binary *Indicator matrix of a set of nominal variables.*

Description

Constructs the indicator matrix for a nominal variables matrix.

Usage

```
NominalMatrix2Binary(Y)
```

Arguments

Y	A matrix with nominal variables measured for a set of individuals. Input must be a matrix with integer values.

Value

G	The binary indicator matrix asociated to the nominal matrix

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

See Also

Nominal2Binary

Examples

```
data(HairColor)
NominalMatrix2Binary(data.matrix(HairColor))
```

PCoA *Principal Coordinates Analysis*

Description

This function calculates principal coordinates analysis using a distante matrix among a set of objets.

Usage

```
PCoA(dis, r = 2)
```

Arguments

dis	Distance matrix between a set ob objects.
r	Number of dimensions for the solution.

Value

An object with has some components:

EigenValues	Eigenvalues of the inner products matrix
Inertia	Variance (Inertia) accounted for each dimension
RowCoordinates	Coordinates for the rows in the reduced space
RowQualities	Qualities of representation of the objects. Squared cosines between the points (vectors) in the full space and the points in the reduced space. Values near 1 indicate good quality

Author(s)

Jose Luis Vicente-Villardon,Julio Cesar Hernandez Sanchez

Maintainer: Jose Luis Vicente-Villardon <villardon@usal.es>

References

Gower,J.C. (1966) *Some distance properties of latent root and vector methods used in multivariate analysis*. Biometrika, 53, 325–338.

See Also

NominalDistances

Examples

```
data(HairColor)
dis = NominalDistances(data.matrix(HairColor))
PCoA(dis,2)
```

PhD_nomCyL *Data set from Survey on Human Resources in Science and Technology carried out by Spanish Statistical Office.*

Description

The sample data corresponds 681 answers from PhD holders, corresponding to people that in 2006 had a doctoral degree and with their residence in Castilla-Le\'on region in Spain. The data come from Survey on Careers of doctorate holders(CDH) carried out by Spanish Statistical Office in 2008.

Usage

```
data(PhD_nomCyL)
```

Format

This data frame contains 681 observation for the following 7 columns:

MS Marital Status:(1:M(Married),2:MLR(Living in a marriage-like relationship), 3:SD (Separated or Divorced),4:SW(Widowed or Single)

SECT Sector of employment(1:BES(Business Enterprise Sector), 2:GS (Government Sector),3:HES(Higher Education Sector), 4:PNP(Private Non Profit))

MIN Minimum education level required for the principal job: (1:mPD(Postdoc),2:mARQ(Advanced Research Qualification), 3:mPG(Post-graduate),4:mGL(Graduate or lower)

DES Desirable education level required for the principal job: (1:dPD(Postdoc),2:dARQ(Advanced Research Qualification), 3:dPG(Post-graduate),4:dGL(Graduate or lower)

PJREL Is your principal job related to your advanced research qualification degree:(1:H(High),2:M(Medium),3:L(Low))

FOSAT Field of science and technology (1:NS(Natural Sciences),2:ET(Engineering and technology), 3:MH(Medical and health sciences),4:AS(Agricultural sciences), 5:SS(Social Sciences),6:H(Humanities))

SOF Principal source of financial support during your research studies (1:F(Fellowship),2:T(Teaching),3:OE (Other Employment),4:R(Reimbursement) ,5:LPSO (LoanPersonalSavingsOther)

Source

Spanish Statistical Office (Survey on Human Resources in Science and Technology, 2006): http://www.ine.es/prodyser/micro_

Examples

```
data(PhD_nomCyL)
```

plot.nominal.logistic.biplot
Graphical representation of a Nominal Logistic Biplot.

Description

Plotting a Nominal Logistic Biplot. There are parameters related to the way in which the biplot is plotted. All the posible parameters have default values

Usage

```
## S3 method for class 'nominal.logistic.biplot'
## S3 method for class 'nominal.logistic.biplot'
plot(x, planex = 1, planey = 2,
QuitNotPredicted = TRUE, ReestimateInFocusPlane = TRUE,
proofMode = FALSE, AtLeastR2 = 0.01, xlimi = -1.5, xlimu = 1.5,
ylimi = -1.5, ylimu = 1.5, linesVoronoi = FALSE, ShowAxis = TRUE,
PlotVars = TRUE, PlotInd = TRUE, LabelVar = TRUE, LabelInd = TRUE,
CexInd = NULL, CexVar = NULL, ColorInd = NULL, ColorVar = NULL,
SmartLabels = FALSE, PchInd = NULL, PchVar = NULL,
LabelValuesVar = NULL, ShowResults = FALSE,...)
```

Arguments

`x`	An object of the class nominal.logistic.biplot.
`planex`	Dimension for X axis.
`planey`	Dimension for Y axis.
`QuitNotPredicted`	Should the non-predicted categories be represented on the graph?
`ReestimateInFocusPlane`	Shuld the item parameters be reestimated using only the dimensiona of the plot.? If FALSE the values of the parameters for other dimensions are set to 0. Default is FALSE
`proofMode`	Should each variable be plotted on a separate plot? If FALSE, a single plot with a legend for identifying each variable is made.
`AtLeastR2`	It establishes the cutting value to plot a variable attending to its Nagelkerke R^2 value. A variable is plotted if its R^2 is higher than this value.
`xlimi`	Minimum value on the x-axis.
`xlimu`	Maximum value on the x-axis.
`ylimi`	Minimum value on the y-axis.
`ylimu`	Maximum value on the y-axis.
`linesVoronoi`	Should the tesselation be plotted.? Default is FALSE and only the category points are plotted for a better reading of the plot.

ShowAxis	Should the axis be shown?
PlotVars	Should the variables (items) be pplotted?
PlotInd	Should the individuals be plotted?
LabelVar	Should the variable labels be shown?
LabelInd	Should the individual labels be shown?
CexInd	Size of the individual points. It can be an array with the cex information for each row.
CexVar	Size of the category points. It can be an array with the cex information for each variable.
ColorInd	Color of the individual points. It can be an array with the color information for each row.
ColorVar	Color for the variables. It can be an array with the color information for each variable.
SmartLabels	Should the text labels be printed accordind to its position on the plot?.
PchInd	Symbol for the individuals. It can be an array with the pch information for each row.
PchVar	Symbol for the variables. It could be an array with the pch information for each variable.
LabelValuesVar	List with the text labels for all the variables. If NULL, initial labels are used.
ShowResults	Should the results of the proccess of calculating the prediction regions be shown?
...	Additional parameters to plot.

Details

The function without parameters plots the nominal.logistic.biplot object with labels in the original data and default values for colors, symbols and sizes for points and lines. Other values of colors, symbols and sizes can be supplied. A single value applies to all the points but an array with different values can be used to improve the undestanding of the plot.-

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

See Also

NominalLogisticBiplot

Examples

```
data(HairColor)
nlbo = NominalLogisticBiplot(HairColor,sFormula=NULL,
numFactors=2,method="EM",penalization=0.2,show=FALSE)
plot(nlbo,QuitNotPredicted=TRUE,ReestimateInFocusPlane=TRUE,
planex = 1,planey = 2,proofMode=TRUE,LabelInd=TRUE,
```

```
AtLeastR2 = 0.01,xlimi=-1.5,xlimu=1.5,ylimi=-1.5,
ylimu=1.5,linesVoronoi = TRUE,SmartLabels = FALSE,
PlotInd=TRUE,CexInd = c(0.6,0.7,0.5,0.4,0.5,0.6,0.7)
,PchInd = c(1,2,3,4,5,6,7),ColorInd="black",PlotVars=TRUE,
LabelVar = TRUE,PchVar = c(1,2,3,4,5),
ColorVar = c("red","black","yellow","blue","green")
,ShowResults=TRUE)
```

---

plotNominalFittedVariable

Function for plotting in the reduced space an unordered and fitted categorical variable.

Description

Graphical representation of a polytomous unordered variable previously fitted in the reduced space, according to the Nominal Logistic Biplot theory. It can be choosen some parameters related to the way in which the variable is plotted.

Usage

```
plotNominalFittedVariable(nameVar, numcateg, beta, varstudyC, rowCoords,
levelsVar = NULL, numFactors = 2, planex = 1, planey = 2, xi = -3.5, xu = 3.5,
yi = -3.5, yu = 3.5, CexVar = 0.7,ColorVar = "blue", PchVar = 0.7,
addToPlot = FALSE, QuitNotPredicted = TRUE, ShowResults = TRUE,
linesVoronoi = TRUE, LabelVar = TRUE)
```

Arguments

`nameVar`	Name of the variable to be plotted.
`numcateg`	Number of categories of the variable.
`beta`	Estimated coefficients matrix.
`varstudyC`	Values of the categorical variable to be plotted. It should be a factor with information about a nominal variable, i.e., an unordered variable.
`rowCoords`	Estimation coordinates for the individuals in the spanned space.
`levelsVar`	Vector with the labels for each level of the variable.
`numFactors`	Dimension of the reduced space.
`planex`	Dimension for X axis.
`planey`	Dimension for Y axis.
`xi`	Minimum value on the x-axis.
`xu`	Maximum value on the x-axis.
`yi`	Minimum value on the y-axis.
`yu`	Maximum value on the y-axis.
`CexVar`	Size of the category points.

`ColorVar`	Color for the variable.
`PchVar`	Symbol for the variable.
`addToPlot`	Should the graph be added to an existing representation?
`QuitNotPredicted`	Should the non-predicted categories be represented on the graph?
`ShowResults`	Should the results of the proccess of calculating the prediction regions be shown?
`linesVoronoi`	Should the tesselation be plotted.? Default is FALSE and only the category points are plotted for a better reading of the plot
`LabelVar`	Should the variable labels be shown?

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

See Also

polylogist

Examples

```
data(Env)
nlbo = NominalLogisticBiplot(Env,sFormula=NULL,
numFactors=2,method="EM",penalization=0.2,show=FALSE)
nameVar = nlbo$dataSet$ColumNames[1]
numcateg = 4
beta = nlbo$VariableModels[,1]$beta
Nagelkerke = nlbo$VariableModels[,1]$Nagelkerke
varstudyC = as.matrix(as.numeric(Env[,1]))
rowCoords = nlbo$RowsCoords
levelsVar = c("M1","M2","M4","M5")
plotNominalFittedVariable(nameVar,numcateg,beta,varstudyC,rowCoords,levelsVar=NULL,
            numFactors=2,planex = 1,planey = 2,xi=-3.5,xu=3.5,yi=-3.5,yu=3.5,
            CexVar=0.7,ColorVar="blue",PchVar=0.7,addToPlot=FALSE,
            QuitNotPredicted=TRUE,ShowResults=TRUE,linesVoronoi=TRUE,LabelVar=TRUE)
```

`plotNominalVariable` *Function for plotting in the reduced space an unordered categorical variable.*

Description

Graphical representation of a polytomous unordered variable in the reduced space, according to the Nominal Logistic Biplot theory. Inside the function, the estimations needed for the variable will be done. It can be choosen some parameters related to the way in which the variable is plotted.

Usage

```
plotNominalVariable(nameVar, nominalVar, estimRows, planex = 1, planey = 2,
xi = -3.5, xu = 3.5, yi = -3.5, yu = 3.5, CexVar = 0.7, ColorVar = "blue",
PchVar = 0.7, addToPlot = FALSE, QuitNotPredicted = TRUE, ShowResults = FALSE,
linesVoronoi = TRUE, LabelVar = TRUE, tol = 1e-04, maxiter = 100,
penalization = 0.1, showIter = FALSE)
```

Arguments

`nameVar`	Name of the variable to be plotted.
`nominalVar`	Values of the categorical variable to be plotted. It should be a factor with information about a nominal variable, i.e., a variable without ordered values.
`estimRows`	Estimation coordinates for the individuals in the spanned space.
`planex`	Dimension for X axis.
`planey`	Dimension for Y axis.
`xi`	Minimum value on the x-axis.
`xu`	Maximum value on the x-axis.
`yi`	Minimum value on the y-axis.
`yu`	Maximum value on the y-axis.
`CexVar`	Size of the category points.
`ColorVar`	Color for the variable.
`PchVar`	Symbol for the variable.
`addToPlot`	Should the graph be added to an existing representation?
`QuitNotPredicted`	Should the non-predicted categories be represented on the graph?
`ShowResults`	Should the results of the proccess of calculating the prediction regions be shown?
`linesVoronoi`	Should the tesselation be plotted.? Default is FALSE and only the category points are plotted for a better reading of the plot
`LabelVar`	Should the variable labels be shown?
`tol`	Value to stop the process of iterations.
`maxiter`	Maximum number of iterations in the process of solving the regression coefficients.
`penalization`	Penalization used in the diagonal matrix to avoid singularities.
`showIter`	Boolean parameter to show the information about the iterations.

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

See Also

`polylogist`

Examples

```
data(HairColor)
nlbo = NominalLogisticBiplot(HairColor,sFormula=NULL,
numFactors=2,method="EM",penalization=0.2,show=FALSE)
nameVar = nlbo$dataSet$ColumNames[2]
nominalVar = HairColor[,2]
estimRows = nlbo$RowsCoords
plotNominalVariable(nameVar,nominalVar,estimRows,planex = 1,planey = 2,
    xi=-1.5,xu=1.5,yi=-1.5,yu=1.5,CexVar=0.7,ColorVar="blue",PchVar=0.7,
    addToPlot=FALSE,QuitNotPredicted=TRUE,ShowResults=TRUE,
    linesVoronoi=TRUE,LabelVar=TRUE,tol = 1e-04, maxiter = 100,
    penalization = 0.3,showIter = FALSE)
```

---

polylogist *Multinomial logistic regression with ridge penalization*

Description

This function does a logistic regression between a dependent variable y and some independent variables x, and solves the separation problem in this type of regression using ridge regression and penalization.

Usage

```
polylogist(y, x, penalization = 0.2, cte = TRUE, tol = 1e-04, maxiter = 200, show = FALSE)
```

Arguments

y	Dependent variable.
x	A matrix with the independent variables.
penalization	Penalization used in the diagonal matrix to avoid singularities.
cte	Should the model have a constant?
tol	Tolerance for the iterations.
maxiter	Maximum number of iterations.
show	Should the iteration history be printed?.

Details

The problem of the existence of the estimators in logistic regression can be seen in Albert (1984), a solution for the binary case, based on the Firth's method, Firth (1993) is proposed by Heinze(2002). The extension to nominal logistic model was made by Bull (2002). All the procedures were initially developed to remove the bias but work well to avoid the problem of separation. Here we have chosen a simpler solution based on ridge estimators for logistic regression Cessie(1992).

Rather than maximizing $L_j(\mathbf{G}|\,\mathbf{b}_{j0}, \mathbf{B}_j)$ we maximize

$$L_j(\mathbf{G}|\,\mathbf{b}_{j0}, \mathbf{B}_j) - \lambda\,(\|\mathbf{b}_{j0}\| + \|\mathbf{B}_j\|)$$

Changing the values of λ we obtain slightly different solutions not affected by the separation problem.

Value

An object of class "polylogist". This has components

`fitted`	Matrix with the fitted probabilities
`cov`	Covariance matrix among the estimates
`Y`	Indicator matrix for the dependent variable
`beta`	Estimated coefficients for the multinomial logistic regression
`stderr`	Standard error of the estimates
`logLik`	Logarithm of the likelihood
`Deviance`	Deviance of the model
`AIC`	Akaike information criterion indicator
`BIC`	Bayesian information criterion indicator

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

References

Albert,A. & Anderson,J.A. (1984),*On the existence of maximum likelihood estimates in logistic regression models*, Biometrika 71(1), 1–10.

Bull, S.B., Mak, C. & Greenwood, C.M. (2002), *A modified score function for multinomial logistic regression*, Computational Statistics and dada Analysis 39, 57–74.

Firth, D.(1993), *Bias reduction of maximum likelihood estimates*, Biometrika 80(1), 27–38

Heinze, G. & Schemper, M. (2002), *A solution to the problem of separation in logistic regression*, Statistics in Medicine 21, 2109–2419

Le Cessie, S. & Van Houwelingen, J. (1992), *Ridge estimators in logistic regression*, Applied Statistics 41(1), 191–201.

Examples

```
data(HairColor)
data = data.matrix(HairColor)
G = NominalMatrix2Binary(data)
mca=afc(G,dim=2)
depVar = data[,1]
nomreg = polylogist(depVar,mca$RowCoordinates[,1:2],penalization=0.1)
nomreg
```

`RidgeMultinomialRegression` *Ridge Multinomial Logistic Regression*

Description

Function that calculates an object with the fitted multinomial logistic regression for a nominal variable. It compares with the null model, so that we will be able to compare which model fits better the variable.

Usage

```
RidgeMultinomialRegression(y, x, penalization = 0.2,
cte = TRUE, tol = 1e-04, maxiter = 200, showIter = FALSE)
```

Arguments

`y`	Dependent variable.
`x`	A matrix with the independent variables.
`penalization`	Penalization used in the diagonal matrix to avoid singularities.
`cte`	Should the model have a constant?
`tol`	Value to stop the process of iterations.
`maxiter`	Maximum number of iterations.
`showIter`	Should the iteration history be printed?.

Value

An object that has the following components:

`fitted`	Matrix with the fitted probabilities
`cov`	Covariance matrix among the estimates
`Y`	Indicator matrix for the dependent variable
`beta`	Estimated coefficients for the multinomial logistic regression
`stderr`	Standard error of the estimates
`logLik`	Logarithm of the likelihood
`Deviance`	Deviance of the model
`AIC`	Akaike information criterion indicator
`BIC`	Bayesian information criterion indicator
`NullDeviance`	Deviance of the null model
`Difference`	Difference between the two deviance values
`df`	Degrees of freedom
`p`	p-value asociated to the chi-squared estimate

`CoxSnell` Cox and Snell pseudo R squared

`Nagelkerke` Nagelkerke pseudo R squared

`MacFaden` MacFaden pseudo R squared

`PercentCorrect`
Percentage of correct classifications

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

References

Albert,A. & Anderson,J.A. (1984),*On the existence of maximum likelihood estimates in logistic regression models*, Biometrika 71(1), 1–10.

Bull, S.B., Mak, C. & Greenwood, C.M. (2002), *A modified score function for multinomial logistic regression*, Computational Statistics and dada Analysis 39, 57–74.

Firth, D.(1993), *Bias reduction of maximum likelihood estimates*, Biometrika 80(1), 27–38

Heinze, G. & Schemper, M. (2002), *A solution to the problem of separation in logistic regression*, Statistics in Medicine 21, 2109–2419

Le Cessie, S. & Van Houwelingen, J. (1992), *Ridge estimators in logistic regression*, Applied Statistics 41(1), 191–201.

See Also

polylogist

Examples

```
data(HairColor)
data = data.matrix(HairColor)
G = NominalMatrix2Binary(data)
mca=afc(G,dim=2)
depVar = data[,1]
rmr = RidgeMultinomialRegression(depVar,mca$RowCoordinates[,1:2],penalization=0.1)
rmr
```

summary.nominal.logistic.biplot

Summary Method Function for Objects of Class 'nominal.logistic.biplot'

Description

This function shows a summary of the principal results for the estimation for individuals and variables, like some Pseudo R-squared indices, the correct classification percentage of each regression, the logLikelihood and "Estimate coefficients", "Std. Error", "z value" or "Pr(>|z|)" values.

Usage

```
## S3 method for class 'nominal.logistic.biplot'
## S3 method for class 'nominal.logistic.biplot'
summary(object,completeEstim,coorInd,...)
```

Arguments

`object`	This parameter keeps the nominal logistic biplot object.
`completeEstim`	Boolean parameter to choose if the estimated coefficients will be printed on screen.Default value is FALSE.
`coorInd`	Boolean parameter to choose if the individual coordinates will be printed on screen.Default value is FALSE.
`...`	Additional parameters to summary.

Details

This function is a method for the generic function summary() for class "nominal.logistic.biplot". It can be invoked by calling summary(x) for an object x of the appropriate class.

Author(s)

Julio Cesar Hernandez Sanchez, Jose Luis Vicente-Villardon

Maintainer: Julio Cesar Hernandez Sanchez <juliocesar_avila@usal.es>

See Also

NominalLogisticBiplot

Examples

```
data(HairColor)
nlbo = NominalLogisticBiplot(HairColor,sFormula=NULL,
numFactors=2,method="EM",penalization=0.2,show=FALSE)
summary(nlbo)
```

Index