

# Package ‘pcpattL’

August 28, 2007

**Type** Package

**Title** Generates design matrix for analysing Likert type items using a loglinear paired comparison model..

**Version** 0.5.3

**Date** 2007-08-28

**Author** Reinhold Hatzinger

**Maintainer** Reinhold Hatzinger <reinhold.hatzinger@wu-wien.ac.at>

**Depends** stats

**Description** Generates design matrix for analysing Likert type items using a loglinear paired comparison model.

**License** GPL 2.0

## R topics documented:

issp2000 . . . . .	1
pcpattL . . . . .	2
xmpl . . . . .	7

<b>Index</b>	<b>8</b>
--------------	----------

---

issp2000	<i>ISSP 2000 Survey on Environmental Issues</i>
----------	---

---

## Description

In 2000 the International Social Survey Programme (ISSP) has addressed the topic of attitudes to environmental protection and preferred government measures for environmental protection. This dataset focusses on six items (with a 5-point Likert type response scale) where respondents from Austria and Great Britain were asked about their perception of environmental dangers.

## Usage

```
data(issp2000)
```

### Format

A data frame with 1595 observations on the following 11 variables. The first six variables are items to be answered on a 5-point Likert type scale with response categories: (1) *extremely dangerous for the environment* to (5) *not dangerous at all for the environment*.

**CAR** air pollution caused by cars

**IND** air pollution caused by industry

**FARM** pesticides and chemicals used in farming

**WATER** pollution of country's rivers, lakes and streams

**TEMP** a rise in the world's temperature

**GENE** modifying the genes of certain crops

**SEX** gender: (1) *male*, (2) *female*

**URB** location of residence: (1) *urban area*, (2) *suburbs of large cities, small town, county seat* (3) *rural area*

**AGE** age: (1) *< 40 years*, (2) *41-59 years*, (3) *60+ years*

**CNTRY** country: (1) *Great Britain*, (2) *Austria*

**EDU** education: (1) *below A-level/matric*, (2) *A-level/matric or higher*

### Source

Dataset: International Social Survey Programme 2000: Environment II (ISSP 2000)

Identification Number: 3440

Central Archive for Empirical Social Research

[http://www.gesis.org/en/data\\_service/issp/index.htm](http://www.gesis.org/en/data_service/issp/index.htm)

Usage regulations:

[http://www.gesis.org/en/data\\_service/order/usage\\_regulations.htm](http://www.gesis.org/en/data_service/order/usage_regulations.htm)

### References

Dittrich, R., Francis, B.J., Hatzinger R., Katzenbeisser, W. (2007), A Paired Comparison Approach for the Analysis of Sets of Likert Scale Responses. *Statistical Modelling*, 7(1):3-28.

### Examples

```
data(issp2000)
str(issp2000)
```

---

pcpattL

*Likert to Paired Comparison Patterns - Design Matrix*

---

### Description

The function `pcpattL` converts a set of likert type responses measured on a common scale into paired comparison patterns, returning a new data frame containing the design matrix for a loglinear paired comparison model. Additionally, the frequencies of these patterns are computed and are stored in the first column of the data frame. Optionally, the function provides all necessary structures (commands, data/design files) to fit the loglinear paired comparisons pattern model in GLIM, which is often more efficient at fitting large loglinear models of this type.

## Usage

```
pcpattL(ctrl, dfr = NULL)
```

## Arguments

ctrl	a list of control parameters If <code>ctrl</code> is not correctly specified an error message is printed.
dfr	a dataframe conforming to the same specifications as the input data file (see below). The default is <code>NULL</code> , i.e., no dataframe is supplied in the call of <code>pcpattL</code> with the data instead supplied through the <code>datafile</code> element in the <code>ctrl</code> list.

## Details

Prior to the call of `pcpattL` the user has to provide a control list `ctrl` (see **The Control List**) and data (either in the form of a dataframe or an external file) which has to conform to a certain structure (see **Input Data**).

The typical usage is

```
desmat <- pcpattL(ctrl)
```

or just `pcpattL(ctrl)` if only the GLIM output is wanted.

The function `pcpattL` allows for different scenarios mainly concerning

- **responses.** The responses to the Likert type items are transformed to paired comparison responses by calculating the difference between each pair of the Likert items. These differences are then reduced to three response categories (*preferred – undecided – not preferred*). The set of paired comparison responses represents a response pattern.
- **item covariates.** The design matrix for the basic model has columns for the items (objects) and for undecided comparisons (i.e., there is one dummy variable for each comparison to reflect if the corresponding item responses are equal and 0 otherwise). Additionally, covariates for two way interaction between comparisons (i.e., for effects resulting from the dependency between two comparisons that have one item in common) can be obtained by setting `blnIntcovs = TRUE`.
- **subject covariates.** For modelling different preference scales for the items according to characteristics of the respondents categorical subject covariates can be included in the design. The corresponding variables are defined as numerical vectors where the levels are specified with consecutive integers starting with 1. This format must be used in the input data file and is also used in all outputs.
- **GLIM output.** If the user specifies `blnGLIMcmds = TRUE` two files are generated one of which contains all GLIM commands to fit a basic loglinear paired comparisons pattern model. The other contains the design matrix optionally including subject covariates. If dependency covariates are requested they are written to a third file. (Please note that the corresponding part of design matrix is transposed in the interactions output file to allow GLIM for using the `$array` facility in case of a large number of parameters to be estimated)

## Value

The output is a dataframe. Each row represents a unique response pattern. If subject covariates are specified, each row instead represents a particular combination of a unique covariate combination with a response pattern. All possible combinations are generated.

The first column contains the counts for the paired comparison response patterns and is labelled with  $Y$ . The next columns are the covariates for the items and the undecided category effects (one for each comparison). These are labelled  $U_{ab}$ , where  $ab$  denotes the comparison between items  $a$  and  $b$ . Optionally, covariates for dependencies between comparisons follow. The columns are labelled  $I_{a.bc}$  denoting the interaction of the comparisons between items  $(a,b)$  and  $(a,c)$  where the common item is  $a$ . If subject covariates are present they are in the rightmost columns and defined to be factors.

Alternatively, the function `pcpattL` does not produce output in  $R$  if GLIM output is requested via `blnGLIMcmds = TRUE`. The output is then written to the corresponding files (see **The Control List** below).

### The Control List

The argument `ctrl` is a list with elements described below. It must be defined prior to the call of `pcpattL`.

**datafile** If the data is to be read from a datafile, the path/name of the datafile to be read

**nitems** the number of Likert items in the datafile

**blnRevert** If the Likert type responses are such that low values correspond to high agreement and high values to low agreement (e.g., (1) *I strongly agree* ... (5) *I strongly disagree*) then `blnRevert` should be specified to be `FALSE`. Otherwise set `blnRevert = TRUE`.

**blnIntcovs** generates covariates for interactions between comparisons if `blnIntcovs = TRUE`.

**cov.sel** a character vector with the names of the subject covariates in the data file to be included into the design matrix. (example: `cov.sel = c("SEX", "AGE")`). If all covariates are to be included the specification can be abbreviated to `cov.sel = "ALL"`. For no covariates specify `cov.sel = ""`.

**blnGLIMcmds** `TRUE`, if GLIM output is wanted. If `blnGLIMcmds = FALSE` the following items can be set to any value (such as a null text string) and are ignored. Please note that if `blnGLIMcmds` is set to be `TRUE` there is no output in  $R$  but instead goes to the the following files.

**glimCmdFile** name of the output file which will contain all necessary commands to fit a basic model (defining all structures and reading the necessary data).

**outFile** name of the data/design file to be read into GLIM. It consists of the response frequencies and the covariates for the objects, the undecided comparison responses and the subject effects.

**intFile** name of the design file for the interaction effects. This file is only generated if `blnIntcovs = TRUE`.

### Input Data

Input data is specified either through an external file (as specified through `datafile` in `ctrl`) or through a dataframe via the argument `dfr`. The input data file if specified must be a plain text file with variable names in the first row as readable via the command `read.table(datafile, header = TRUE)`. The leftmost columns must be the responses to the Likert items optionally followed by columns for categorical subject covariates. These have to be specified such that the categories are represented by consecutive integers starting with 1. Missing values are treated such that rows with one or more NAs are removed from the data and a message is printed. For an example see [xmpl](#) or the file `xmpl.dat` in the package's `data/` directory.

**Warning**

Care has to be taken in case of a larger number of Likert items ( $\geq 7$ ) and/or of covariate level combinations. The tractable size of the design matrix depends on the working memory available to R.

**Author(s)**

Reinhold Hatzinger

**References**

Dittrich, R., Francis, B.J., Hatzinger R., Katzenbeisser, W. (2007), A Paired Comparison Approach for the Analysis of Sets of Likert Scale Responses. *Statistical Modelling*, 7(1):3-28.

**Examples**

```
## Not run:

## EXAMPLE 1: TYPICAL USAGE
## not run because input data file does not exist

# defining the ctrl list
# would be typically read from file using source()

testex1<-list(
  datafile      = "test/test.dat",
  nititems     = 5,
  blnRevert    = FALSE,
  blnIntcovs   = FALSE,

  cov.sel      = c("SEX", "URB"),

  blnGLIMcmds  = TRUE,
  glimCmdFile  = "test/test.gli",
  outFile      = "test/test.design",
  intFile      = ""           # since blnIntcovs = FALSE
)

# call

pcpattL(testex1)

## End(Not run)

## EXAMPLE 2: WITH LOADED DATAFRAME

data(xmpl)    # example data in package

testex2<-list(
  datafile     = "",          # dataframe used
  nititems    = 3,
  blnRevert   = FALSE,
  blnIntcovs  = TRUE,

  cov.sel     = "ALL",
```

```

    blnGLIMcmds = FALSE,    # no GLIM output
    glimCmdFile = "",
    outFile     = "",
    intFile     = ""
  )

dsgnmat <- pcpattL(testex2, xmpl)

print(head(dsgnmat))

## EXAMPLE 3: ILLUSTRATING THE ISSP2000 EXAMPLE
##   simplified version of the analysis as
##   given in Dittrich et.al.(2007)

data(issp2000)

testex3<-list(
  datafile      = "",
  nitems       = 6,
  blnRevert    = FALSE,
  blnIntcovs   = FALSE,

  cov.sel      = c("SEX", "EDU"),

  blnGLIMcmds  = FALSE,
  glimCmdFile  = "",
  outFile      = "",
  intFile      = ""
)

design <- pcpattL(testex3, issp2000)

# - fit null multinomial model (basic model for items without
#   subject covariates) through Poisson distribution.
# - SEX:EDU parameters are nuisance parameters
# - the last item (GENE) becomes a reference item
#   in the model and is aliased; all other items
#   are compared to this last item

# item parameters with undecided effects and no covariate effects.

summary(glm(y~SEX:EDU + CAR+IND+FARM+WATER+TEMP+GENE
            + U12+U13+U23+U14+U24+U34+U15+U25+U35+U45+U16+U26+U36+U46+U56,
            family=poisson, data=design))

# now add main effect of SEX on items

summary(glm(y~SEX:EDU + CAR+IND+FARM+WATER+TEMP+GENE
            + (CAR+IND+FARM+WATER+TEMP+GENE):SEX
            + U12+U13+U23+U14+U24+U34+U15+U25+U35+U45+U16+U26+U36+U46+U56,
            family=poisson, data=design))

```

---

`xmpl`*Example Data Set*

---

**Description**

Data to illustrate the usage of `pcpattL`.

**Usage**

```
data(xmpl)
```

**Format**

A data frame with 100 observations on the following 5 variables.

**I1** responses to first item, (1) *strong agreement*, to (5) *strong disagreement*

**I2** responses to second item, like I1

**I3** responses to thirs item, like I1

**SEX** (1) *male*, (2) *female*

**EDU** (1) *low education*, (2) *high education*

All values are numeric.

**Details**

Datasets and/or dataframes used in `pcpattL` are required to have the follwing structure:

- All values must be numeric.
- The item responses must be in the leftmost columns (such as I1 to I3 above).
- Categorical subject covariates follow the item responses (rightmost columns) and their levels must be specified as consecutive integers. If in a used datafile or dataframe these are defined as **R** factors they will be converted to integers. This is not possible if characters are used as factor levels and, consequently, `pcpattL` will produce an error.

**Examples**

```
data(xmpl)
head(xmpl)
```

# Index

## \*Topic **datasets**

issp2000, 1

xmpl, 7

## \*Topic **misc**

pcpattL, 2

issp2000, 1

pcpattL, 2, 7

xmpl, 4, 7