Package ‘KernSmoothIRT’

March 6, 2013

Type Package

Title Nonparametric Item Response Theory

Version 5.0

Date 2013-01-11

Author Angelo Mazza, Antonio Punzo, Brian McGuire

Maintainer Brian McGuire <mcguirebc@gmail.com>

Description This package fits nonparametric item and option characteristic curves using kernel smoothing. It allows for optimal selection of the smoothing bandwidth using cross-validation and a variety of exploratory plotting tools.

License GPL-2

LazyLoad yes

Depends Rcpp, plotrix, rgl

LinkingTo Rcpp

NeedsCompilation yes

Repository CRAN

Date/Publication 2013-03-06 19:54:07

R topics documented:

KernSmoothIRT-package ........................................... 2
BDI ................................................................. 3
HIV ................................................................. 3
itemcor ............................................................. 5
ksIRT .............................................................. 5
PCA ................................................................. 10
plot.ksIRT ......................................................... 10
KernSmoothIRT-package

Description

This package is meant for the graphical analysis of multiple choice test and questionnaire data within a nonparametric approach. Basically, it performs the fitting of item and options characteristic curves using kernel smoothing techniques and provides a variety of exploratory plots.

Details

Package: KernSmoothIRT
Type: Package
Version: 1.0
Date: 2011-05-29
Requires: plotrix
License: GPL-2
LazyLoad: yes

The function `ksIRT` creates a `ksIRT` object using kernel smoothing. Plotting the `ksIRT` object creates a variety of exploratory plots, including: item characteristic curves, option characteristic curves, density plots, expected value plots, and a probability simplex plot for the top 3 highest probability options of each item.

Author(s)

Angelo Mazza, Antonio Punzo, Brian McGuire

Maintainer: Brian McGuire <mcguirebc@gmail.com>

References

**BDI**

*The Beck Depression Inventory Data*

**Description**

The Beck Depression Inventory (Beck, Rush, Shaw & Emery, 1979) is a 21-question self-report inventory, one of the most widely used instruments for measuring the severity of depression. Each question has a set of four possible answer choices, ranging in intensity. In this dataset, the respondents are 242 McGill University students. The data were collected by Prof. David Zuroff.

**Usage**

```r
data(BDI)
BDIkey
BDIresponses
```

**Format**

- **BDIkey** a numeric vector with the highest level ordinal response for each of the 21 questions.
- **BDIresponses** is a numeric matrix with examinees responses. Each column represents a question and each row represents a subject; thus, **BDIresponses** has 242 rows and 21 columns.

**Source**


**References**


---

**HIV**

*The HIV Data*

**Description**

These data come from the Voluntary HIV Counseling and Testing Efficacy Study (2000a,b), conducted in 1995-1997, by the Center for AIDS Prevention Studies, University of California, San Francisco. This study was concerned with the effectiveness of HIV counseling and testing in reducing risk behavior for the sexual transmission of HIV. As part of this study, respondents were surveyed about their attitude toward condom use via a bank of 15 items. Respondents were asked how much they agreed with each of the statements on a 4-point response scale.
Usage

data(HIV)
HIV
HIVkey

Format

SITE the site the participant attended. "Ken" for Kenya, "Tan" for Tanzania and "Tri" for Trinidad.

GENDER gender of participant

AGE age (at last birthday) of participant

I1 Using condoms is a good way to protect you and your sex partner from diseases people can get from having sex.

I2 Sex doesn’t feel as good when you use a condom.

I3 It’s embarrassing to buy condoms.

I4 Using condoms is a good way to prevent an unwanted pregnancy.

I5 It’s embarrassing to put on a condom (put a condom on a man).

I6 Most of your friends think you should use condoms whenever you have sex, including with your spouse.

I7 Condoms often break or slip off during sex.

I8 If your sex partner wants to use a condom, I’d suspect he/she may be having sex with someone else.

I9 Most of your friends use condoms when they have sex with a new partner.

I10 It is easy to buy condoms in the area where I now live.

I11 Most of your friends think that condoms are uncomfortable.

I12 Most of your friends think you should always use condoms when having sex with a new person.

I13 It is easy to get condoms in my home town without having to pay for them.

I14 Condoms cost too much.

I15 Most people your age are using condoms these days.

Details

Given the "negative" wording of the items I2, I3, I5, I7, I8, I11, and I14, in order to uniform the data, their scores has been reversed.

Source

The whole data set can be downloaded from http://caps.ucsf.edu/research/datasets/
References


**itemcor**

*ksIRT* - *kernel smoothing in Item Response Theory*

**Description**

This function returns a vector of point polyserial correlations for each item.

**Usage**

```r
itemcor(x)
```

**Arguments**

- `x` a `ksIRT` object to be analyzed.

**References**


---

**ksIRT**

*ksIRT* - *kernel smoothing in Item Response Theory*

**Description**

This function fits nonparametric item and options characteristic curves using kernel smoothing techniques. Within the `KernSmoothIRT` package, it provides the relevant data for the graphical analysis of multiple choice test and questionnaire data.
Usage

ksIRT(responses, key, format, kernel = c("gaussian","quadratic","uniform"), itemlabels, weights, miss = "option", bandwidth = c("Silverman", "CV"), RankFun = "sum", SubRank, thetadist = list("norm",/zero.noslash,1), groups = FALSE)

## S3 method for class 'ksIRT'
print(x,...)

Arguments

responses    input data matrix with options selected from each individual for each item. Rows represent individuals, columns represent items. Alternatively, a data.frame or list can be specified. Missing values are inserted as NA.

key          an optional numeric vector. If provided key must be a numeric vector with the number of entries equal to the number of items. If the items are multiple choice, key should contain the option that corresponds to the correct response. If the data are rating-scale, key should contain the largest option value for each item. In this case, the weight assigned to each option is equal to its option number. More complicated weighting schemes, such as partial credit, can be specified in the weights argument. If weights is specified, key must be left blank.

format       a numeric scalar or vector specifying the type of items. If all of the items are multiple choice, then format = 1. If all of the items are rating-scale or partial credit, then format = 2. If all of the items are nominal items, then format = 3. If the test has a mixture of items of different formats, then format is a vector with length equal to the number of items with entries of 1 for each multiple choice item and 2 for each rating-scale item. For more complicated weighting schemes use the weights argument.

kernel       a character string specifying the kernel function. kernel must be either "gaussian", "quadratic" or "uniform". The default is "gaussian".

itemlabels   optional list of labels for each item. If omitted, each item will be labelled according to its numerical order. These labels will be used in plotting.

weights      optional list that may be used in lieu of including key. Specifying weights allows for more complicated weighting schemes than the default. Its length must be equal to the number of items and each entry must be a matrix with option numbers in the first row and option weights in the second row. If weights is omitted and format=1, then weights are given according to key. If weights is omitted and format=2, then an option weight equals the option number is given to each response. If weights is omitted and format=3, then weights are set to zero.

miss         a character string specifying the method used to manage missing responses. The default value, miss="option", considers the missing responses as a further option, labeled as NA, with zero weight. Such NA option will be added to the plot of the Option Characteristic Curves. Alternatively, a different weight for the NA option may be specified through the NAweight argument. miss="random.unif" substitutes NAs with options randomly chosen from the possible ones for the corresponding item.
miss="random.multinom" does the same substitution as miss="random.unif" but each option has a probability of being selected proportional to its relative frequency.

miss="omit" excludes from the analysis all the subjects with at least one omitted response.

NAweight a scalar value that specifies the weight given to missing responses when miss="option". The default is zero.

evalpoints an optional numeric vector that specifies the quantiles at which to estimate the Option Characteristic Curves. If unspecified, the default is nevalpoints evenly spaced values with end points determined according to the number of subjects and the distribution specified with the thetadist argument.

nevalpoints an optional scalar value that specifies the number of evenly spaced points at which curves are estimated. This value is used as an alternative to a user defined vector in the evalpoints argument. The default value is 51. The end points are determined according to the number of subjects and to the distribution specified for the thetadist argument. If both nevalpoints and evalpoints are specified, then evalpoints takes precedence.

bandwidth either "Silverman", "CV" or a numeric vector specifying, for each item, the bandwidth to use for kernel smoothing. The default value, bandwidth="Silverman", is a numeric vector computed following the well-known Silverman's rule of thumb. If bandwidth="CV", then the bandwidth is chosen for each item through cross-validation.

RankFun a function that is used to rank subjects. The default value is "sum". Another common choice is "mean".

SubRank an optional numeric vector specifying the rank of each of the subjects. If unspecified, subjects will be ranked according to the function passed through the argument RankFun.

thetadist a list specifying the distribution to be used to thetadist (see Ramsay, 1991, p. 615) the subjects. By default a standard normal distribution is used. A different distribution can be adopted by specifying the first element of the list as "norm", "beta", "unif", "gamma", etc. where the character string is the same as used in the subjscoresummary function qnorm(), qbeta(), qunif(), qgamma(). The other elements of the list should be the distribution parameters as required by the subjscoresummary function chosen.

groups an optional vector of length equal to the number of subjects containing the group designation of each subject. Adding this option allows for comparisons between groups using the Differential Item Functioning tools (see details section).

x a ksIRT object to be printed.

... further parameters

Details

When bandwidth="Silverman", the rule of thumb of Silverman (1986, p. 45) is implemented with the formula: \( 1.06 \times \sigma_h \times n \times (-0.2) \), where \( n \) is the number of subjects and \( \sigma_h \) is the standard deviation of the subjscoresummary associated to the subjects according
to the distribution specified with \texttt{thetadist}. Note that when \texttt{thetadist}=\texttt{list("norm", mean, sd)}, \texttt{sigma.hat} is the value specified for \texttt{sd}.

Printing the \texttt{ksIRT} object shows the point polyserial correlation correlation between each item and the overall test score.

\textbf{Value}

Returned from this function is a \texttt{ksIRT} object which is a list with the following components:

- \texttt{nitem}: an integer indicating the number of items.
- \texttt{nsubj}: an integer indicating the number of subjects.
- \texttt{nevalpoints}: an integer indicating the number of points for curve estimation.
- \texttt{binaryresp}: a matrix of binary responses. Each row corresponds to a single option. The first three columns specify the item, the option, and the corresponding weight. Each additional column is a binary indicator of whether a subject selected that option.
- \texttt{OCC}: a matrix with the first 3 columns the same as \texttt{binaryresp} and an additional column for each quantile at which the option characteristic curves have been estimated. The additional columns contain the kernel smoothed probabilities of selecting each option.
- \texttt{stderrs}: a matrix as \texttt{OCC} containing the standard errors of \texttt{OCC}.
- \texttt{subjscore}: a vector containing the observed score of each subject.
- \texttt{itemlabels}: a list containing the label for each item.
- \texttt{thetadist}: a list indicating the distribution used to rank subjects (see \texttt{thetadist} in Arguments).
- \texttt{subjtheta}: a vector of quantile ranks for each subject on the distribution specified in \texttt{thetadist}.
- \texttt{evalpoints}: a vector with the \texttt{subjscore} used in curve estimation.
- \texttt{subjscoresummary}: a vector of \texttt{subjscore}, of probabilities \(.05, .25, .50, .75, .95\), for the observed overall scores.
- \texttt{subjscoresummaryevalpoints}: a vector as \texttt{subjscore} but computed on \texttt{subjtheta}.
- \texttt{SmthWgts}: a matrix containing the kernel weights.
- \texttt{scale}: a vector indicating whether each item is multiple-choice, rating-scale or nominal; 1 indicates multiple-choice, 0 indicates rating-scale, 3 indicates nominal.
- \texttt{format}: returns the \texttt{format} argument passed at function call.
- \texttt{bandwidth}: a vector containing the bandwidths for each item.
- \texttt{DIF}: a list of \texttt{ksIRT} objects created for each of the subgroups specified by \texttt{groups}.
- \texttt{groups}: returns the \texttt{groups} argument passed at function call.
- \texttt{itemcor}: a vector containing the point polyserial correlation for each item.
- \texttt{RCC}: a list of \texttt{nsubj} vectors containing the normalized likelihood for each value in \texttt{evalpoints}.
- \texttt{subjthetaML}: the maximum likelihood estimate for the expected total score of each subject.
References


Examples

```r
## Psych1/zero.noslash1 data
data(Psych101)
Psych1 <- ksIRT(responses = Psychresponses[1:100,], key = Psychkey, format = 1)
Psych1

plot(Psych1, plottype="OCC", item=c(24, 25, 92, 96))
plot(Psych1, plottype="EIS", item=c(24, 25, 92, 96))
plot(Psych1, plottype="tetrahedron", items=c(24, 92))
plot(Psych1, plottype="triangle", items=c(24, 92))
plot(Psych1, plottype="PCA")
plot(Psych1, plottype="RCC", subjects=c(33, 92))

PCA(Psych1)
subjEIS(Psych1)
subjETS(Psych1)
subjOCC(Psych1, stype="ObsScore")
bsubjscore(Psych1)
bsubjthetaML(Psych1)

plot(Psych1, plottype="expected")
plot(Psych1, plottype="sd")
plot(Psych1, plottype="density")

## HIV data
data(HIV)
HIVsubset <- HIV[c(c(1:5),c(1508:1558),c(2934:2984)),]
gr2 <- as.character(HIVsubset$SITE)
DIF2 <- ksIRT(res=HIVsubset[,-(1:3)], key=HIVkey, format = 2, groups=gr2, miss="omit")

plot(DIF2, plottype="expectedDIF", lwd=2)
plot(DIF2, plottype="densityDIF", lwd=2)
plot(DIF2, plottype="EISDIF", item=c(6,11))

### Ordinal Survey Data
data(BDI)
BDI1 <- ksIRT(responses=BDIresponses, key=BDIkey, format = 2, miss="omit")

plot(BDI1, plottype="OCC", items=1:4)
plot(BDI1, plottype="sd")
plot(BDI1, plottype="density", ylim=c(0,0.1))
```
PCA

*ksIRT* - kernel smoothing in Item Response Theory

**Description**
This function returns a vector with each subject's expected test score.

**Usage**
```r
PCA(x, ...)  
```

**Arguments**
- `x`: a *ksIRT* object to be analyzed.
- `...`: other arguments to pass to the `prcomp` function from the `stats` package.

**References**


---

**plot.ksIRT**

*Plot Method for ksIRT* - kernel smoothing in Item Response Theory

**Description**
The plot method for *ksIRT* objects includes a variety of exploratory plotting tools.

**Usage**
```r
## S3 method for class 'ksIRT'  
plot(x, plottype = c("OCC", "EIS", "density", "expected", "sd", "triangle", "tetrahedron", "RCC", "EISDIF", "OCCDIF", ...), items = "all", ...)  
```

**Arguments**
- `x`: a *ksIRT* created with `ksIRT()`
- `plottype`: the type of plot to be created (see section Details below). With the default value, `plottype="OCC"`, option characteristic curves are plotted.
- `items`: a vector containing the items to be plotted. With the default value, `items="all"`, all items are plotted.
**alpha**
either alpha=FALSE or a scalar indicating the confidence level to be used for creating confidence intervals. It is used with plottype="EIS", and the default value is alpha=.05, or with plottype="OCC", and the default value is alpha=FALSE.

**subjects**
a vector specifying the subjects to plot. This argument is only used when plottype="RCC".

**axistype**
a character string specifying the display variable to be used on the x-axis. The default is axistype="distribution", which uses the subjscoresummary of the distribution specified in thetadist. The alternative is axistype="scores" which displays the expected score.

**main, xlab, ylab, xlim, ylim, cex**
plotting parameters (see plot()) with useful defaults.

**...**
further plotting parameters.

**Details**

Possible values for plottype are:

- **plottype="density"** produces a simple kernel density plot of the observed scores.
- **plottype="EIS"** plot of the expected item scores for each of the item numbers in the items argument.
- **plottype="OCC"** plot of the option characteristic curves for each of the item numbers in the items argument.
- **plottype="expected"** plot of the observed vs. expected scores.
- **plottype="sd"** plot of the standard deviation of observed scores.
- **plottype="RCC"** plots the RCC and actual score for each subject specified by the subjects argument.
- **plottype="triangle"** produces a triangle simplex plot with the highest 3 probability options for each item specified by the items argument.
- **plottype="tetrahedron"** produces a tetrahedron simplex plot with the highest 4 probability options for each item specified by the items argument. The tetrahedron plot can be rotated by using the mouse.
- **plottype="PCA"** produces Principle Component Analysis plot of the test.

Below are values for plottype used for Differential Item Functioning (DIF) plots. They are available only if the groups argument is specified when creating the ksIRT object:

- **plottype="densityDIF"** plots density of observed scores for each of the different groups.
- **plottype="expectedDIF"** plots pairwise expected value comparison plots for each of the different groups.
- **plottype="EISDIF"** plots expected item scores for each of the different groups. Accepts the same arguments as plottype="EIS", but by default does not show confidence intervals. This can be changed with the alpha argument.
- **plottype="OCCDIF"** plots option characteristic curves for each of the different groups. Accepts the same arguments as plottype="OCC"
Value

No values are returned from the plot function.

References


Description

These data come from a multiple-choice examination given to 379 students in an introductory course in psychology in the Christmas exam period of 1989 at McGill. The test consisted of 100 multiple-choice items, each having four response options.

Usage

data(Psych1/zero.noslash1)
Psychkey
Psychresponses

Format

Psychkey is a numeric vector with the correct responses for each of the 100 questions on the exam.
Psychresponses is a numeric matrix with students’ answers. Each column represents a question and each row represents a student.

Details

See ksIRT() help page for more information and examples on how to use these data with the KernSmoothIRT package.

Source


References

subjEIS

ksIRT - kernel smoothing in Item Response Theory

Description

This function returns a matrix containing each subject’s expected item score. The rows represent items and the columns, subjects.

Usage

subjEIS(x)

Arguments

x

a ksIRT object to be analyzed.

References


subjEISDIF

ksIRT - kernel smoothing in Item Response Theory

Description

This function applies only when the groups argument is specified. It returns a matrix containing each subject’s expected item score. The rows represent items and the columns, subjects.

Usage

subjEISDIF(x)

Arguments

x

a ksIRT object to be analyzed.

References


subjETS

*ksIRT* - kernel smoothing in *Item Response Theory*

**Description**

This function returns a vector with each subject's expected test score.

**Usage**

`subjETS(x)`

**Arguments**

- `x`  
  a *ksIRT* object to be analyzed.

**References**


subjETSDIF

*ksIRT* - kernel smoothing in *Item Response Theory*

**Description**

This function applies only when the `groups` argument is specified. It returns a vector with each subject's expected test score.

**Usage**

`subjETSDIF(x)`

**Arguments**

- `x`  
  a *ksIRT* object to be analyzed.

**References**


Description

This function returns a list containing a matrix for each item. Each matrix in the list contains a row for each option with each column representing a subject with the probability of selecting that option for each subject.

Usage

subjOCC(x, stype = c("ObsScore","ExpectedScore","MLScore","Theta","MLTheta"))

Arguments

x a ksIRT object to be analyzed.
stype the scale on which to evaluate each subject. stype = "ObsScore" uses the subject's observed test score. stype = "ExpectedScore" uses the subject's expected test score. stype = "MLScore" uses the maximum likelihood estimate for the subject's overall score. stype = "Theta" uses the subject's rank on the thetadist scale. stype = "MLTheta" uses the maximum likelihood estimate for the subject on the thetadist scale.

References


Description

This function applies only when the groups argument is specified. It returns a list containing a matrix for each item for each of the different groups. Each matrix in the list contains a row for each option with each column representing a subject with the probability of selecting that option for each subject.

Usage

subjOCCDIF(x, stype = c("ObsScore","Theta","ThetaML","ScoreML"))
Arguments

- **x**: a ksIRT object to be analyzed.
- **stype**: the scale on which to evaluate each subject. stype = "ObsScore" uses the subject's total score. stype = "Theta" uses the subject's rank on the thetadist scale. stype = "ThetaML" uses the maximum likelihood estimate for the subject on the thetadist scale. codestype = "ScoreML" uses the maximum likelihood estimate for the subject on the overall test score scale.

References


---

**subjscore**

*ksIRT::kernel smoothing in Item Response Theory*

Description

This function returns a vector of scores for each subject.

Usage

`subjscore(x)`

Arguments

- **x**: a ksIRT object to be analyzed.

References


### subjscoreML

**Description**

This function returns the maximum likelihood estimate for each subject.

**Usage**

```r
subjscoreML(x)
```

**Arguments**

- `x`  
  a `ksIRT` object to be analyzed.

**References**


---

### subjthetaML

**Description**

This function returns the maximum likelihood estimate for each subject.

**Usage**

```r
subjthetaML(x)
```

**Arguments**

- `x`  
  a `ksIRT` object to be analyzed.

**References**


Index

*Topic **datasets**
  BDI, 3
  HIV, 3
  Psych1/zero.noslash1, 12

BDI, 3
BDIkey (BDI), 3
BDIresponses (BDI), 3

HIV, 3
HIVkey (HIV), 3

itemcor, 5

KernSmoothIRT (KernSmoothIRT-package), 2
KernSmoothIRT-package, 2
ksIRT, 5

PCA, 10
plot (plot.ksIRT), 10
plot.ksIRT, 10
print.ksIRT (ksIRT), 5
Psych1/101, 12
Psychkey (Psych1/101), 12
Psychresponses (Psych1/101), 12

subjEIS, 13
subjEISDIF, 13
subjETS, 14
subjETSDIF, 14
subjOCC, 15
subjOCCDIF, 15
subjscore, 16
subjscoreML, 17
subjthetaML, 17