Package ‘regDIF’

June 22, 2021

Title  Regularized Differential Item Functioning
Version  1.0.0
Author  William Belzak
Maintainer  William Belzak <wbelzak@gmail.com>
License  MIT + file LICENSE
Encoding  UTF-8
LazyData  true
RoxygenNote  7.1.1
VignetteBuilder  knitr
Imports  stats (> 3.0.0), utils, statmod
Suggests  knitr, testthat (>= 2.1.0), covr, rmarkdown
NeedsCompilation  yes
Repository  CRAN
Date/Publication  2021-06-22 07:30:02 UTC

R topics documented:

regDIF-package  .................................................. 2
coef.regDIF  .................................................. 2
ida  .................................................. 3
plot.regDIF  .................................................. 3
print.regDIF  .................................................. 4
regDIF  .................................................. 4
summary.regDIF  .................................................. 7

Index  8
### Description

Regularized Differential Item Functioning

### Details

**regDIF** is a package that performs regularization of differential item functioning (DIF) in item response theory (IRT) and confirmatory factor analysis (CFA) models using a penalized expectation-maximization algorithm.

### Author(s)

William Belzak <wbelzak@gmail.com>

### coef.regDIF

#### Description

Coefficient function for regDIF function

#### Usage

```r
## S3 method for class 'regDIF'
coef(object, tau = NULL, method = "bic", ...)
```

#### Arguments

- **object**: Fitted regDIF model object.
- **tau**: Optional character or numeric indicating the tau(s) at which the model coefficients are returned. For character value, may be "tau.min", which returns model coefficients for the value of tau at which the minimum fit statistic is identified. For numeric, the value(s) provided corresponds to the value(s) of tau.
- **method**: Character value indicating the model fit statistic to be used for determining "tau.min". Default is "bic". May also be "aic".
- **...**: Additional arguments to be passed through to coef.

#### Value

NULL
**ida**

*Simulated data example with multiple DIF covariates*

**Description**

A simulated dataset containing six binary items and three DIF covariates

**Usage**

ida

**Format**

A data frame with 500 rows and 9 variables:

- item1
- item2
- item3
- item4
- item5
- item6
- age
- gender
- study ...

**plot.regDIF**

*Plot function for regDIF function*

**Description**

Plot function for regDIF function

**Usage**

```r
## S3 method for class 'regDIF'
plot(x, y = NULL, method = "bic", legend.seed = 123, ...)
```

**Arguments**

- `x` Fitted regDIF model object.
- `y` Unused for plotting regDIF model object.
- `method` Fit statistic to use for identifying DIF effects in plot.
- `legend.seed` Random seed to sample line colors and line types for DIF effects in plot.
- `...` Additional arguments to be passed through to `plot`.
Value

a "plot" object for a "regDIF" fit

print.regDIF

Print function for regDIF function

Description

Print function for regDIF function

Usage

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

Arguments

x  Fitted regDIF model object.
... Additional arguments to be passed through print.

Value

NULL

regDIF

Regularized Differential Item Functioning

Description

Identify DIF in item response theory models using regularization.

Usage

regDIF(item.data,
pred.data,
prox.data = NULL,
item.type = NULL,
pen.type = NULL,
tau = NULL,
num.tau = 100,
alpha = 1,
gamma = 3,
anchor = NULL,
stdz = TRUE,
control = list())
**Arguments**

- `item.data`: Matrix or data frame of item responses. See below for supported item types.
- `pred.data`: Matrix or data frame of predictors affecting item responses (DIF) and latent variable (impact). See `control` option below to specify different predictors for impact model.
- `prox.data`: Optional vector of observed scores to serve as a proxy for the latent variable. If a vector is supplied, a multivariate regression model will be fit to the data. The default is `NULL`, indicating that latent scores will be estimated during model estimation.
- `item.type`: Optional character value or vector indicating the type of item to be modeled. The default is `NULL`, corresponding to a 2PL or graded item type. Different item types may be specified for a single model by providing a vector equal in length to the number of items in `item.data`. The options include:
  - "Rasch" - Slopes constrained to 1 and intercepts freely estimated.
  - "2PL" - Slopes and intercepts freely estimated.
  - "Graded" - Slopes, intercepts, and thresholds freely estimated.
- `pen.type`: Optional character value indicating the penalty function to use. The default is `NULL`, corresponding to the LASSO function. The options include:
  - "lasso" - The least absolute selection and shrinkage operator (LASSO), which controls DIF selection through $\tau$ (tau).
  - "mcp" - The minimax concave penalty (MCP), which controls DIF selection through $\tau$ (tau) and estimator bias through $\gamma$ (gamma). Uses the firm-thresholding penalty function.
  - "grp.lasso" - The group version of the LASSO penalty, which selects intercept and slope DIF effects on each background characteristic together.
  - "grp.mcp" - The group version of the MCP function.
- `tau`: Optional numeric vector of tau values $\geq 0$. If tau is supplied, this overrides the automatic construction of tau values. Must be non-negative and in descending order, from largest to smallest values (e.g., `seq(1,0,-.01)`).
- `num.tau`: Numeric value indicating how many tau values to fit. The default is 100.
- `alpha`: Numeric value indicating the alpha parameter in the elastic net penalty function. Alpha controls the degree to which LASSO or ridge is used during regularization. The default is 1, which is equivalent to LASSO. NOTE: If using MCP penalty, alpha may not be exactly 0.
- `gamma`: Numeric value indicating the gamma parameter in the MCP function. Gamma controls the degree of tapering of DIF effects as tau decreases. Larger gamma leads to faster tapering (less bias but possibly more unstable optimization), whereas smaller gamma leads to slower tapering (more bias but more stable optimization). Default is 3. Must be greater than 1.
- `anchor`: Optional numeric value or vector indicating which item response(s) are anchors (e.g., `anchor = 1`). Default is `NULL`, meaning at least one DIF effect per covariate will be fixed to zero as tau approaches 0 (required to identify the model).
- `stdz`: Logical value indicating whether to standardize DIF and impact predictors for regularization. Default is `TRUE`, as it is recommended that all predictors be on the same scale.
Optional list of different model specifications and optimization parameters. May be:

- **impact.mean.data**: Matrix or data frame of predictors, which allows for a different set of predictors to affect the mean impact equation compared to the item response DIF equations. Default includes all predictors from pred.data.

- **impact.var.data**: Matrix or data frame with predictors for variance impact. See above. Default includes all predictors in pred.data.

- **tol**: Convergence threshold of EM algorithm. Default is $10^{-5}$.

- **maxit**: Maximum number of EM iterations. Default is 5000.

- **adapt.quad**: Logical value indicating whether to use adaptive quadrature to approximate the latent variable. The default is FALSE. NOTE: Adaptive quadrature is not supported yet.

- **num.quad**: Numeric value indicating the number of quadrature points to be used. For fixed-point quadrature, the default is 21 points when all item responses are binary or else 51 points if at least one item is ordered categorical.

- **optim.method**: Character value indicating which optimization method to use. Default is "MNR", which updates the impact and item parameter estimates using Multivariate Newton-Raphson. Another option is "UNR", which updates estimates one-at-a-time using univariate Newton-Raphson, or a single iteration of coordinate descent. A third option is "CD", or coordinate descent with complete iterations through all parameters until convergence. "MNR" will be faster in most cases, although "UNR" may achieve faster results when the number of predictors is large.

- **start.values**: List of numbers assigned as starting values to the regDIF procedure. List must contain only the following names: impact, for mean and variance impact parameters, in the order that is given by an object of class coef.regDIF; base, for base intercept and slope parameters, in order given by a coef.regDIF object; and finally, dif, for intercept and slope DIF parameters, again in order given by a coef.regDIF object.

### Value

Function returns an object of class `regDIF`, which is a list of results from the regularization routine.

### Examples

```r
library(regDIF)
head(ida)
item.data <- ida[,1:6]
pred.data <- ida[,7:9]
prox.data <- rowSums(item.data)
fit <- regDIF(item.data, pred.data, prox.data, num.tau = 10)
summary(fit)
```
Summary function for regDIF function

Description
Summary function for regDIF function

Usage

```r
## S3 method for class 'regDIF'
summary(object, method = "bic", ...)
```

Arguments

- `object`: Fitted regDIF model object.
- `method`: Fit statistic to use for displaying minimum tau model.
- `...`: Additional arguments to be passed through `summary`.

Value

`NULL`
Index

* datasets
  ida, 3
* package
  regDIF-package, 2

coef.regDIF, 2
ida, 3
plot.regDIF, 3
print.regDIF, 4
regDIF, 4
regDIF-package, 2
summary.regDIF, 7