

Package ‘microEnem’

December 27, 2023

Title Manipulação dos Microdados do Enem

Version 0.0.0.9000

Author Alexandre Jaloto [aut, cre] (<<https://orcid.org/0000-0002-5291-1768>>)

Description Manipulação dos microdados do Exame Nacional do Ensino Médio (Enem). Calcula a nota na escala oficial do Enem.

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Imports data.table,
mirt,
mirtCAT

URL <https://github.com/inepoficial/microEnem>

BugReports <https://github.com/inepoficial/microEnem/issues>

Depends R (>= 2.10)

R topics documented:

abre.resp	1
calc.nota	2
constantes	4
mod.caderno	5

Index	6
--------------	----------

abre.resp	<i>Abre o vetor de resposta</i>
-----------	---------------------------------

Description

Abre o vetor de resposta da variável TX_RESPOSTAS dos microdados

Usage

abre.resp(unico)

Arguments

unico vetor único de respostas de um ou mais sujeitos no formato da variável TX_RESPOSTAS dos microdados do Enem

Value

Um objeto de classe *matrix* em que as colunas são os itens e as linhas são os sujeitos

Examples

```
# importar primeiros 100 casos dos microdados de 2019 (é preciso direcionar para a pasta onde o arquivo se localiza)
micro <- data.table::fread('MICRODADOS_ENEM_2019.csv', nrow = 100)

resp <- abre.resp(micro$TX_RESPOSTAS_LC)
```

calc.nota	<i>Calcula nota do Enem</i>
-----------	-----------------------------

Description

Calcula a nota de um sujeito na escala do Enem (500, 100).

Usage

```
calc.nota(resps, codigo = NULL, lingua = NULL)
```

Arguments

resps vetor de respostas de um ou mais sujeitos. O vetor deve ser no mesmo formato da variável TX_RESPOSTAS dos microdados do Enem.

É importante destacar que desde 2010 os cadernos de Linguagens e Códigos têm cinco itens de língua inglesa e cinco itens de língua espanhola, por isso esses cadernos possuem 50 questões.

Em algumas edições de microdados, o vetor de respostas da variável TX_RESPOSTAS_LC possui 50 caracteres e em outros, 45. Em 2019, o vetor possui 50 caracteres e em 2022, 45. Em ambos os casos, o vetor contém somente a resposta do sujeito à língua estrangeira selecionada no momento da inscrição. No caso do vetor com 50 caracteres, as respostas aos cinco itens da outra língua são marcados com 9. A função *calc.nota* transforma internamente essa resposta 9 em NA e esses itens não são considerados para o cálculo da nota do sujeito.

codigo código da prova disponível no dicionário dos microdados. Essa informação também está disponível no objeto *dic.cad* deste pacote

lingua vetor com indicação da língua estrangeira escolhida pela pessoa. Essa informação corresponde à variável TP_LINGUA nos microdados (0 para inglês, 1 para espanhol). Só é necessário para o ano de 2022.

Details

A nota é calculada pelo método Expected a posteriori (EAP), com 40 pontos de quadratura de -4 a 4. A média da distribuição priori é 0 e o desvio padrão, 1. No Enem as notas estão em uma escala com média 500 e desvio padrão 100. A referência dessa escala são os concluintes regulares de escola pública do Enem 2009. Ou seja, a média desses alunos no Enem 2009 foi 500 e o desvio padrão, 100.

Os parâmetros dos itens estão divulgados em uma escala com média 0 e desvio padrão 1. A referência dessa escala é a amostra utilizada na primeira calibração dos itens do Enem, em 2009. Para posicionar os parâmetros na escala oficial do Enem, aplicamos as seguintes equações de transformação:

$$a_{enem} = \frac{a_{01}}{s}$$

$$b_{enem} = b_{01} * s + m$$

$$c_{enem} = c_{01}$$

Onde a_{enem} , b_{enem} e c_{enem} são os parâmetros dos itens na escala oficial do Enem, a_{01} , b_{01} e c_{01} são os parâmetros dos itens divulgados nos microdados, e k e d são as constantes de transformação da escala dos parâmetros divulgados para a escala oficial. Essas constantes estão disponibilizadas no objeto constantes deste pacote.

Value

as notas na escala oficial do Enem

Examples

```
# importar primeiros 100 casos dos microdados de 2022
# (é preciso direcionar para a pasta onde o arquivo se localiza)
micro <- data.table::fread('D:/Microdados/Enem/MICRODADOS_ENEM_2022.csv', nrows = 100)

# selecionar os casos do caderno 1065 (LC)
resp <- subset(micro, micro$CO_PROVA_LC == 1065)

# calcular a nota
nota <- calc.nota(resps = resp$TX_RESPOSTAS_LC, codigo = 1065, lingua = resp$TP_LINGUA)
nota

# comparar com a nota oficial
all.equal(resp$NU_NOTA_LC, nota)

# calcular a nota de um sujeito que não está nos microdados
# vetor de resposta de uma pessoa fictícia que respondeu o caderno 1065 de LC em 2022 (45 caracteres)
resp <- c('BBDABBDABAADCBABBADAACBDDDDDEACACBCACAABBBECBEC')

# calcular a nota
nota <- calc.nota(resp, codigo = 1065, lingua = 0)
nota

# vetor de resposta de uma pessoa fictícia que respondeu o caderno 511 de LC em 2019 (50 caracteres)
resp <- c('99999BBDABBDABAADCBABBADAACBDDDDDEACACBCACAABBBECBEC')

# calcular a nota
nota <- calc.nota(resp, codigo = 511)
nota
```

constantes

*Constantes de transformação***Description**

Constantes de transformação da escala dos itens divulgados nos microdados para a escala oficial do Enem

Usage

constantes

Format

Um objeto do tipo `data.frame` com as constantes de cada área.

Temos duas escalas: a escala oficial do Enem, cuja referência são os concluintes regulares de escolas públicas do Enem 2009 (média 500, desvio padrão 100); e a escala dos itens divulgados nos microdados (escala dos microdados), cuja referência é a amostra de calibração do Enem 2009 (média 0, desvio padrão 1).

Para calcular as constantes de transformação de uma escala para a outra, aplicamos a equalização linear (Hambleton et al., 1991). Inicialmente estimamos a nota dos primeiros 300.000 sujeitos do banco dos microdados em cada área, de acordo com as especificações em `calc.nota`. Em seguida, estabelecemos a igualdade entre a padronização das notas desses sujeitos:

$$\frac{Y_i - \bar{Y}}{DP_y} = \frac{X_i - \bar{X}}{DP_x}$$

onde Y_i representa a nota do sujeito i na escala oficial do Enem e \bar{Y} e DP_y representam a média e o desvio padrão das notas dessa amostra de 300.000 sujeitos nessa mesma escala. X_i representa a nota do sujeito i na escala dos microdados. \bar{X} e DP_x representam a média e o desvio padrão das notas da amostra nessa escala.

Nessa equação, consideramos que as notas padronizadas das duas escalas são iguais, pois provêm da mesma amostra. Se isolarmos Y_i , teremos (Muñiz, 1997):

$$Y_i = \frac{DP_y}{DP_x} X_i + \bar{Y} - \frac{DP_y}{DP_x} \bar{X}$$

Dessa equação, extraímos as constantes k e d :

$$k = \frac{DP_y}{DP_x}$$

$$d = \bar{Y} - \frac{DP_y}{DP_x} \bar{X}$$

$$d = \bar{Y} - k\bar{X}$$

Portanto, para transformar a nota da escala dos microdados para a escala oficial do Enem, utilizamos a seguinte equação:

$$Y_i = kX_i + d$$

`mod.caderno`*Modelo mirt do caderno*

Description

Gera um modelo do pacote mirt para um caderno do Enem

Usage

```
mod.caderno(codigo = NULL)
```

Arguments

<code>codigo</code>	código da prova disponível no dicionário dos microdados. Essa informação também está disponível no objeto <code>dic.cad</code> deste pacote
---------------------	---

Value

Um objeto mirt com os parâmetros dos itens. Para informações sobre esse objeto, consulte a documentação do pacote mirt.

Examples

```
mod <- mod.caderno(511)

# parâmetros dos itens
mirt::coef(mod, IRTpars = TRUE, simplify = TRUE)

# gráfico com informação do item 1
item <- mirt::extract.item(mod, 1)
theta <- matrix(seq(-4,4, by = .1))
info <- mirt::iteminfo(item, theta)
plot(theta, info, type = 'l', main = 'Informação do item 1')
```

Index

* **datasets**

constantes, [4](#)

abre.resp, [1](#)

calc.nota, [2](#)

constantes, [4](#)

mod.caderno, [5](#)