Package 'extidy'

December 13, 2022

Title Preprocess the data prior to analysis.

Version 1.0.0

Description Preprocess the data prior to analysis.

License GPL (>= 3)

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.2

Imports httr,vroom,readxl

NeedsCompilation no

Author bin Wang [aut, cph, cre]

Maintainer bin Wang <binwang@pku.edu.cn>

R topics documented:

DelMiss *Delete variables with missing values*

Description

Whether to delete missing variables with low variance. The default option is "yes". If skipped, it may result in failure during modeling.

Usage

```
DelMiss(PID, OutPath = "default")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.

Value

An R6 class object containing the variable(s) without missing values.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = DelMiss(PID=res$PID)
    FuncExit(PID = res$PID)
```

DelNearZeroVar *Delete variables with low variance*

Description

Whether to delete variables with low variance. The default option is "yes". If skipped, it may result in failure to build models.

Usage

```
DelNearZeroVar(PID, OutPath = "default")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.

Value

An R6 class object containing the variable(s) with acceptable variance.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = DelNearZeroVar(PID=res$PID)
    FuncExit(PID = res$PID)
```

FuncExit *End the module analysis*

Description

End the module analysis

Usage

```
FuncExit(PID)
```

Arguments

PID	chr. Program ID. It should be the same with the PID generated by initial functions.

Value

Exit status

Author(s)

Bin Wang (corresponding author)

Examples

```
res = InitTidy()
   res1 = LoadTidy(PID=res$PID, UseExample="example#1")
   res2 = DelNearZeroVar(PID=res$PID)
   FuncExit(PID = res$PID)
```

InitTidy *Initialize ExpoTidy module*

Description

Initialize ExpoTidy module analysis. It can generate an R6 class object.

Usage

```
InitTidy()
```

Details

It is designed to tidy the data for the target model analysis.

Value

An R6 class object.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
   FuncExit(PID = res$PID)
```

LoadTidy *Load data file for ExpoTidy module*

Description

Load data file for ExpoCros module

Usage

```
LoadTidy(PID, UseExample = "default", DataPath=NULL, VocaPath=NULL)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by ExpoCros
UseExample	chr. Method of uploading data. If "default",user should upload their own data files, or use "example#1" provided by this module.
DataPath	chr. Input directory of data file, e.g. "D:/test/eg_data_biolink.xlsx". It should be noted that the slash symbol is "/", not "\".
VocaPath	chr. Input directory of vocabulary file, e.g. "D:/test/eg_voca_biolink.xlsx". It should be noted that the slash symbol is "/", not "\".

Value

An R6 class object containing the input data.

Author(s)

Bin Wang

Examples

```
res <- InitTidy()
   res = LoadTidy(PID = res$PID, UseExample = "example#1")
   FuncExit(PID = res$PID)
```

`TransClass` *Classify variables into various groups*

Description

Classify variables into various groups

Usage

```
TransClass(PID, OutPath = "default", Group, Vars, LevelTo)
```

Arguments

`PID`	chr. Program ID. It must be the same with the PID generated by initial functions.
`OutPath`	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
`Group`	lgl. Whether to separate dataset into train and test data for processing data.
`Vars`	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. Users can also choose available variables by selecting "Other" option, and copy the variables by clicking "Available vars". It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
`LevelTo`	The number of levels to convert variables to.

Value

An R6 class object containing the variable(s) after classifying data into various levels.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransClass(PID=res$PID, Group= FALSE, Vars="X1", LevelTo="4")
    FuncExit(PID = res$PID)
```

TransDistr *Transform variable distribution*

Description

Transform variable distribution

Usage

```
TransDistr(PID,OutPath = "default", Vars, Method)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Vars	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. Users can also choose available variables by selecting "Other" option, and copy the variables by clicking "Available vars". It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
Method	chr. Methods used for imputation. Available options include "lod" or "cart". For "lod" method, limit of detection (LOD) should be included in the "Vocabulary" file.

Value

An R6 class object containing the variable(s) after transforming distribution.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransDistr(PID=res$PID, Vars="X6,X7", Method="log10")
    FuncExit(PID = res$PID)
```

TransDummy *Transform factor variables into dummy ones*

Description

Transform factor variables into dummy ones

Usage

```
TransDummy(PID, OutPath = "default", Vars="default")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Vars	chr. Variables to be transformed as dummy variables. It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3". If "default", all the factor variables will be transformed into dummy ones. These variables need to be transformed as factor ones in previous transform step using TransType function.

Value

An R6 class object containing the variable(s) after transforming the factor variables into dummy ones.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransDummy(PID=res$PID, Vars="default")
    FuncExit(PID = res$PID)
```

TransGroup *Transform exposure groups*

Description

Transform exposure groups

Usage

```
TransGroup(PID, OutPath = "default", Vars="default", ToGroup)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Vars	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. Users can also choose available variables by selecting "Other" option, and copy the variables by clicking "Available vars". It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
ToGroup	chr. Label the group of the target variables. Four common used names are recommended, including Exposure, Metabolome, Proteome, and Immunome. Users can also label the groups as you like.

Value

An R6 class object containing the variable(s) after grouping the variables.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransGroup(PID=res$PID, Vars="X4,X5", ToGroup = "G1")
    FuncExit(PID = res$PID)
```

TransImput *Missing data imputation.*

Description

Missing data imputation.

Usage

```
TransImput(PID,OutPath = "default",Group,Vars,Method)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Group	lgl. Whether to separate dataset into train and test data for processing data.
Vars	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. Users can also choose available variables by selecting "Other" option, and copy the variables by clicking "Available vars". It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
Method	Methods used for imputation. Available options include "lod" or "cart" methods. For "lod" method, limit of detection (LOD) should be included in the "Vocabulary" file.

Value

An R6 class object containing variable(s) with imputation.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransImput(PID=res$PID, Group=TRUE, Vars="all.x", Method="lod")
    FuncExit(PID = res$PID)
```

TransScale *Scale variables*

Description

Scale variables

Usage

```
TransScale(PID,OutPath = "default", Group = T, Vars, Method = "normal",
    Direct="positive", RangeLow="0", RangeUpper="1")
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Group	lgl. T (or TRUE) and F (or FALSE). Whether to separate dataset into train and test data for processing data.
Vars	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. Users can also choose available variables by selecting "Other" option, and copy the variables by clicking "Available vars". It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
Method	chr. Scaling methods. Available options include "normal" and "range".
Direct	chr. Direction to be transformed, Available options include "positive" and "negative".
RangeLow	num. Lower limit for range method.
RangeUpper	num. Upper limit for range method. It should be greater than the lower limit.

Value

An R6 class object containing the variable(s) after scaling data.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransScale(PID=res$PID, Group= TRUE, Vars="all.x", Method="normal")
    FuncExit(PID = res$PID)
```

TransType *Transform data type*

Description

Transform data type

Usage

```
TransType(PID,OutPath = "default",Vars,To)
```

Arguments

PID	chr. Program ID. It must be the same with the PID generated by initial functions.
OutPath	chr. Output file directory, e.g. "D:/test". It should be noted that the slash symbol is "/", not "\". If "default", the current working directory will be set.
Vars	Variables to be imputed. Available options include: "all.x", all exposure variables; "all.c", all covariates; "all.cx", combination of All X and All C. It should be noted that there is fixed format for the entering characters separated with comma and without space, e.g., "X1,X2,X3".
To	chr. Indicate the type of the chosen variables to be transformed into. Available options include "integer", "numeric", "character", "factor", "logical", and "date".

Value

An R6 class object containing the variable(s) after transforming data type.

Author(s)

Bin Wang

Examples

```
res = InitTidy()
    res1 = LoadTidy(PID=res$PID, UseExample="example#1")
    res2 = TransType(PID=res$PID, Vars = "X1,X2", To = "character")
    FuncExit(PID = res$PID)
```

Index