

Package ‘InterVA5’

May 11, 2018

Type Package

Title Replicate and Analyse 'InterVA5'

Version 1.0.1

Date 2018-04-19

Author Jason Thomas, Zehang Li, Tyler McCormick, Sam Clark, Peter Byass

Maintainer Jason Thomas <jarathomas@gmail.com>

Description Provides an R version of the 'InterVA5' software (<<http://www.interva.net>>) for coding cause of death from verbal autopsies. It also provides simple graphical representation of individual and population level statistics.

License GPL-3

NeedsCompilation no

Repository CRAN

RoxygenNote 6.0.1

Date/Publication 2018-05-11 19:49:24 UTC

R topics documented:

InterVA5-package	2
causetextV5	2
COMCAT.interVA5	3
CSMF.interVA5	4
CSMF5	5
DataCheck5	6
InterVA5	7
InterVA5.plot	10
print.interVA5_summary	11
probbaseV5	11
RandomVA5	12
summary.interVA5	12

Index	14
--------------	-----------

InterVA5-package	<i>Perform InterVA5 algorithm and provide graphical summarization of COD distribution.</i>
------------------	--

Description

Computes individual cause of death and population cause-specific mortality fractions using the InterVA5 algorithm. Provides a simple graphical representation of the result.

Details

To get the most up-to-date version of the package, as well as the past versions, please check the github repository at: <https://github.com/jarathomas/InterVA5-R-Replicate/>

Package: InterVA5
 Type: Package
 Version: 1.0
 Date: 2018-02-01
 License: GPL-3

Author(s)

Jason Thomas, Zehang Li, Tyler McCormick, Sam Clark
 Maintainer: Jason Thomas <jarathomas@gmail.com>

References

<http://www.interva.net/>

causetextV5	<i>Translation list of COD codes</i>
-------------	--------------------------------------

Description

This is the translation of COD abbreviation codes into their corresponding full names.

Format

A data frame with the translation of codes to their names for 3 pregnancy statuses, 61 CODs (both the version of COD only and COD with group code), and 6 circumstances of mortality (COMCAT).

Examples

```
data(causetextV5)
```

COMCAT.interVA5	<i>Summarize population level mortality fraction by Circumstance of Mortality Category</i>
-----------------	--

Description

The function takes input of a list of va object and calculates the mortality fraction by Circumstance of Mortality Category.

Usage

```
COMCAT.interVA5(va)
```

Arguments

va	The list of va object to summarize.
----	-------------------------------------

Value

dist.cod	The cause-specific mortality fraction (including undetermined category).
----------	--

Author(s)

Jason Thomas, Zehang LI, Tyler McCormick, Sam Clark

See Also

[CSMF5](#)

Examples

```
## Not run:
data(RandomVA5)
# only fit first 20 observations for a quick illustration
RandomVA5 <- RandomVA5[1:20, ]

sample.output <- InterVA5(RandomVA5, HIV = "h", Malaria = "v",
  write=TRUE, directory = tempdir(),
  filename = "VA5_result", output = "extended", append = FALSE)
## Get CSMF without plots
comcat <- COMCAT.interVA5(sample.output$VA5)

## End(Not run)
```

CSMF.interVA5	<i>Summarize population level cause-specific mortality fraction as InterVA5 suggested.</i>
---------------	--

Description

The function takes input of a list of va object and calculates the cause-specific mortality fraction. It only calculates CSMF5 as aggregation of up to the third largest causes.

Usage

```
CSMF.interVA5(va)
```

Arguments

va The list of va object to summarize.

Value

dist.cod The cause-specific mortality fraction (including undetermined category).

Author(s)

Jason Thomas, Zehang LI, Tyler McCormick, Sam Clark

See Also

[CSMF5](#)

Examples

```
## Not run:
data(RandomVA5)
# only fit first 20 observations for a quick illustration
RandomVA5 <- RandomVA5[1:20, ]

sample.output <- InterVA5(RandomVA5, HIV = "h", Malaria = "v", write=TRUE,
  directory = tempdir(), filename = "VA5_result", output = "extended", append = FALSE)
## Get CSMF without plots
csmf <- CSMF.interVA5(sample.output$VA5)

## End(Not run)
```

 CSMF5

Summarize and plot a population level distribution of va probabilities.

Description

The function takes input of a list of va object and produces a summary plot for the population distribution.

Usage

```
CSMF5(va, top.aggregate = NULL, InterVA.rule = FALSE, noplot = FALSE,
      type = "bar", top.plot = 10, min.prob = 0, ...)
```

Arguments

va	The list of va object to summarize.
top.aggregate	Integer indicating how many causes from the top need to go into summary. The rest of the probabilities goes into an extra category "Undetermined". When set to NULL, default is all causes to be considered. This is only used when InterVA.rule set to "FALSE".
InterVA.rule	If it is set to "TRUE", only the top 3 causes reported by InterVA5 is calculated into CSMF as in InterVA5. The rest of probabilities goes into an extra category "Undetermined". Default set to "FALSE".
noplot	A logical value indicating whether the plot will be shown. If it is set to "TRUE", only the CSMF will be returned.
type	An indicator of the type of chart to plot. "pie" for pie chart; "bar" for bar chart.
top.plot	the maximum number of causes to plot in bar plot
min.prob	The minimum probability that is to be plotted in bar chart, or to be labeled in pie chart.
...	Arguments to be passed to/from graphic function barplot , pie , and more graphical paramters (see par). They will affect the main title, size and font of labels, and the radius of the pie chart.

Value

dist.cod	The population probability of CODs.
----------	-------------------------------------

Author(s)

Jason Thomas, Zehang LI, Tyler McCormick, Sam Clark

See Also

[CSMF.interVA5](#)

Examples

```
## Not run:
data(RandomVA5)
# only fit first 20 observations for a quick illustration
RandomVA5 <- RandomVA5[1:20, ]

sample.output <- InterVA5(RandomVA5, HIV = "h", Malaria = "v", write = FALSE,
  directory = tempdir(), filename = "VA5_result", output = "extended",
  append = FALSE)

## Get CSMF by considering only top 3 causes reported by InterVA5.
## This is equivalent to using CSMF.interVA5() command Note that
## it's different from using all top 3 causes, since they may not
## all be reported
CSMF.summary <- CSMF5(sample.output, InterVA.rule = TRUE,
  noplot = TRUE)

## Population level summary using pie chart
CSMF.summary2 <- CSMF5(sample.output, type = "pie",
  min.prob = 0.01, main = "population COD distribution using pie chart",
  clockwise = FALSE, radius = 0.7, cex = 0.7, cex.main = 0.8)

## Population level summary using bar chart
CSMF.summary3 <- CSMF5(sample.output, type = "bar",
  min.prob = 0.01, main = "population COD distribution using bar chart",
  cex.main = 1)
CSMF.summary4 <- CSMF5(sample.output, type = "bar",
  top.plot = 5, main = "Top 5 population COD distribution",
  cex.main = 1)

## End(Not run)
```

DataCheck5

Data cleaning for InterVA-5 algorithm

Description

This function implements the data cleaning steps in the InterVA5 software.

Usage

```
DataCheck5(Input, id, probbaseV5, write)
```

Arguments

Input	original data vector for one observation coded by 0 (absence), 1 (presence), and NA (missing).
-------	--

id	id for this observation
probbaseV5	matrix of probbaseV5
write	logical indicator of writing to file

Value

Output	new data vector
firstPass	message for the first pass check
secondPass	message for the second pass check

Author(s)

Jason Thomas Zehang Li, Tyler McCormick, Sam Clark

References

<http://www.interva.net/>

See Also

[InterVA5.plot](#)

Examples

```
data(RandomVA5)
data(probbaseV5)
probbaseV5 <- as.matrix(probbaseV5)
RandomVA5 <- as.matrix(RandomVA5)
input <- as.character(RandomVA5[1, ])
input[which(toupper(input) == "N")] <- "0"
input[which(toupper(input) == "Y")] <- "1"
input[which(input != "1" & input != "0")] <- NA
input <- as.numeric(input)
output <- DataCheck5(Input=input, id="d1", probbaseV5=probbaseV5, write=TRUE)
```

InterVA5

Provide InterVA5 analysis on the data input.

Description

This function implements the algorithm in the InterVA5 software. It produces individual cause of death (COD) and population cause-specific mortality fractions. The output is saved in a .csv file specified by user. The calculation is based on the conditional and prior distribution of 61 CODs. The function can also save the full probability distribution of each individual to file. All information about each individual is saved to a va class object.

Usage

```
InterVA5(Input, HIV, Malaria, write = TRUE, directory = NULL,
         filename = "VA5_result", output = "classic", append = FALSE,
         groupcode = FALSE, ...)
```

Arguments

Input	A matrix input, or data read from csv files in the same format as required by InterVA5. Sample input is included as data(RandomVA5).
HIV	An indicator of the level of prevalence of HIV. The input should be one of the following: "h"(high), "l"(low), or "v"(very low).
Malaria	An indicator of the level of prevalence of Malaria. The input should be one of the following: "h"(high), "l"(low), or "v"(very low).
write	A logical value indicating whether or not the output (including errors and warnings) will be saved to file. If the value is set to TRUE, the user must also provide a value for the parameter "directory".
directory	The directory to store the output from InterVA5. It should either be an existing valid directory, or a new folder to be created. If no path is given and the parameter for "write" is true, then the function stops and an error message is produced.
filename	The filename the user wish to save the output. No extension needed. The output is in .csv format by default.
output	"classic": The same delimited output format as InterVA5; or "extended": delimited output followed by full distribution of cause of death probability.
append	A logical value indicating whether or not the new output should be appended to the existing file.
groupcode	A logical value indicating whether or not the group code will be included in the output causes.
...	not used

Details

Be careful if the input file does not match InterVA5 input format strictly. The function will run normally as long as the number of symptoms are correct. Any inconsistent symptom names will be printed in console as warning. If there is a wrong match of symptom from warning, please change the input to the correct order.

Value

ID	identifier from batch (input) file
MALPREV	selected malaria prevalence
HIVPREV	selected HIV prevalence
PREGSTAT	most likely pregnancy status
PREGLIK	likelihood of PREGSTAT

PRMAT	likelihood of maternal death
INDET	indeterminate outcome
CAUSE1	most likely cause
LIK1	likelihood of 1st cause
CAUSE2	second likely cause
LIK2	likelihood of 2nd cause
CAUSE3	third likely cause
LIK3	likelihood of 3rd cause
wholeprob	full distribution of causes of death
COMCAT	most likely circumstance of mortality
COMNUM	likelihood of COMCAT
wholeprob	full distribution of causes of death

Author(s)

Jason Thomas Zehang Li, Tyler McCormick, Sam Clark

References

<http://www.interva.net/>

See Also

[InterVA5.plot](#)

Examples

```
data(RandomVA5)
# only fit first 5 observations for a quick illustration
RandomVA5 <- RandomVA5[1:5, ]

## to get easy-to-read version of causes of death make sure the column
## orders match interVA5 standard input this can be monitored by checking
## the warnings of column names

sample.output1 <- InterVA5(RandomVA5, HIV = "h", Malaria = "1", write=FALSE,
  directory = tempdir(), filename = "VA5_result", output = "extended", append = FALSE)

## Not run:
## to get causes of death with group code for further usage
sample.output2 <- InterVA5(RandomVA5, HIV = "h", Malaria = "1",
  write = FALSE, directory = tempdir(), filename = "VA5_result_wt_code", output = "classic",
  append = FALSE, groupcode = TRUE)

## End(Not run)
```

InterVA5.plot	<i>Plot an individual-level distribution of va probabilities.</i>
---------------	---

Description

The function takes an input of a single va object and produces a summary plot for it.

Usage

```
InterVA5.plot(va, type = "bar", min.prob = 0.01, ...)
```

Arguments

va	A va object
type	An indicator of the type of chart to plot. "pie" for pie chart; "bar" for bar chart.
min.prob	The minimum probability that is to be plotted in bar chart, or to be labeled in pie chart.
...	Arguments to be passed to/from graphic function barplot , pie , and more graphical paramters (see par). They will affect the main title, size and font of labels, and the radius of the pie chart.

See Also

[CSMF5](#)

Examples

```
## Not run:
data(RandomVA5)
# only fit first 20 observations for a quick illustration
RandomVA5 <- RandomVA5[1:20, ]
#' sample.output <- InterVA5(RandomVA5, HIV = "h", Malaria = "v", write = FALSE,
  directory = tempdir(), filename = "VA5_result", output = "extended", append = FALSE)

## Individual level summary using pie chart
InterVA5.plot(sample.output$VA5[[3]], type = "pie", min.prob = 0.01,
  main = "1st sample VA analysis using pie chart", clockwise = FALSE,
  radius = 0.6, cex = 0.6, cex.main = 0.8)

## Individual level summary using bar chart
InterVA5.plot(sample.output$VA5[[3]], type = "bar", min.prob = 0.01,
  main = "2nd sample VA analysis using bar chart", cex.main = 0.8)

## End(Not run)
```

```
print.interVA5_summary
```

Print method for summary of the results obtained from InterVA5 algorithm

Description

This function prints the summary message of the fitted results.

Usage

```
## S3 method for class 'interVA5_summary'  
print(x, ...)
```

Arguments

x	summary of InterVA5 results
...	not used

```
probbaseV5
```

Conditional probability of InterVA5

Description

This is the table of conditional probabilities of symptoms given CODs, along with prior probabilities in the first row. The values are from InterVA-5

Format

A data frame with 354 observations on 87 variables. The first row contains observations corresponding to prior probabilities; while the subsequent observations (rows 2 - 354) are the conditional probabilities.

Examples

```
data(probbaseV5)
```

RandomVA5	<i>200 records of Sample Input</i>
-----------	------------------------------------

Description

This is a dataset consisting of 200 arbitrary sample input deaths in the acceptable format of InterVA5. Any dataset that needs to be analyzed by this package should be in the same format. The order of the input fields must not be changed.

Format

200 arbitrary input records.

Examples

```
data(RandomVA5)
```

summary.interVA5	<i>Summary of the results obtained from InterVA5 algorithm</i>
------------------	--

Description

This function prints the summary message of the fitted results.

Usage

```
## S3 method for class 'interVA5'
summary(object, top = 5, id = NULL,
        InterVA.rule = TRUE, ...)
```

Arguments

object	fitted object from InterVA5()
top	number of top CSMF to show
id	the ID of a specific death to show
InterVA.rule	If it is set to "TRUE", only the top 3 causes reported by InterVA5 is calculated into CSMF as in InterVA5. The rest of probabilities goes into an extra category "Undetermined". Default set to "TRUE".
...	not used

References

<http://www.interva.net/>

Examples

```
## Not run:
data(RandomVA5)
# only fit first 20 observations for a quick illustration
RandomVA5 <- RandomVA5[1:20, ]

## to get easy-to-read version of causes of death make sure the column
## orders match interVA5 standard input this can be monitored by checking
## the warnings of column names

sample.output1 <- InterVA5(RandomVA5, HIV = "h", Malaria = "1",
  write = FALSE, directory = tempdir(), filename = "VA5_result",
  output = "extended", append = FALSE)

summary(sample.output1)
summary(sample.output1, top = 10)
summary(sample.output1, id = "sample3")

## End(Not run)
```

Index

*Topic **InterVA**

- DataCheck5, 6
- InterVA5, 7
- InterVA5-package, 2
- InterVA5.plot, 10
- print.interVA5_summary, 11
- summary.interVA5, 12

*Topic **datasets**

- causetextV5, 2
- probbaseV5, 11
- RandomVA5, 12

*Topic **interVA**

- COMCAT.interVA5, 3
- CSMF.interVA5, 4
- CSMF5, 5

barplot, 5, 10

causetextV5, 2
COMCAT.interVA5, 3
CSMF.interVA5, 4, 5
CSMF5, 3, 4, 5, 10

DataCheck5, 6

InterVA5, 7
InterVA5-package, 2
InterVA5.plot, 7, 9, 10

par, 5, 10
pie, 5, 10
print.interVA5_summary, 11
probbaseV5, 11

RandomVA5, 12

summary.interVA5, 12