Package 'IcTemporalPatternDiscovery'

February 6, 2015
Type Package
Title Population-level estimation method that estimates risk by combining a self-controlled and cohort design.
Version 1.0.0
Date 2014-12-22
Author Tomas Bergvall
Maintainer Tomas Bergvall < tomas.bergvall@who-umc.org>
Description HOMEWORK- add more
Depends DatabaseConnector
Imports SqlRender,
plyr
License Apache License 2.0
R topics documented:
appendToIctpdAnalysisList calculateStatisticsIc createIctpdAnalysis getDbIctpdData loadIctpdAnalysisList runIctpd runIctpd writeIctpdAnalysisList 1
Index 1
appendToIctpdAnalysisList
Append to list of ictpdAnalysis
Description
appendToIctpdAnalysisList adds an object of type ictpdAnalysis to a list.
Usage
<pre>appendToIctpdAnalysisList(ictpdAnalysis, ictpdAnalysisList = NULL)</pre>

2 calculateStatisticsIc

Arguments

```
ictpdAnalysis Object to append to the list. ictpdAnalysisList
```

The list to add the object to. If not provided, a new list will be created.

Value

A list of objects of type ictpdAnalysis.

Examples

```
## Not run:
#Constructing a list of analyses:
analysisList <- NULL
analysis <- createIctpdAnalysis(analysisId = 1, censor = TRUE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis)</pre>
analysis <- createIctpdAnalysis(analysisId = 2, censor = FALSE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis, analysisList)</pre>
#You can store the list:
writeIctpdAnalysisListToFile(analysisList,"ictpdAnalysisList.csv")
#And load it back later:
analysisList <- loadIctpdAnalysisList("ictpdAnalysisList.csv")</pre>
#You can run all analyses across a set of exposure-outcome pairs with one command:
connectionDetails <- createConnectionDetails(dbms="sql server", server="server_ip")</pre>
exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),
                                   exposureConceptId = c(1501700, 1545958, 1551803))
results <- runIctpdAnalyses(connectionDetails,</pre>
                             cdmDatabaseSchema = "cdm_schema.dbo",
                             resultsDatabaseSchema = "results.dbo",
                             exposureOutcomePairs = exposureOutcomePairs)),
                             ictpdAnalysisList = ictpdAnalysisList)
## End(Not run)
```

calculateStatisticsIc compute the IC statistics

Description

compute the IC statistics

Usage

```
calculateStatisticsIc(ictpdData, multipleControlPeriods = "110",
  multipleRiskPeriods = "10000", shrinkage = 0.5, icPercentile = 0.025,
  metric = "IC025")
```

createIctpdAnalysis 3

Arguments

ictpdData An object containing the counts, as created using the getDbIctpdData function. multipleControlPeriods

Defines the control periods to use where 100 means the control period defined by controlPeriodStart/End, 010 means the period -30 to -1 day before prescription and 001 means the control period on the day of prescription

multipleRiskPeriods

Defines the risk periods to use 10000 is 1-30 days, 01000 is 1 to 360 days, 00100 is 31 to 90 days, 00010 is 91 to 180 and 00001 is 721 to 1080 days after

prescription default is '10000'

shrinkage used in IRR calculations, required >0 to deal with 0 case counts, but

larger number means more shrinkage. default is 0.5

icPercentile The lower bound of the credibility interval for the IC values (IClow). default is

0.025,

metric Defines wether the output will contain the point estimate or the lower bound.

Available input is 'IC and 'IC025' default is 'IC025'

Value

An object of type ictpdResults containing the results.

Examples

 $create {\tt IctpdAnalysis} \quad \textit{createSccAnalysisDetails}$

Description

createIctpdAnalysis generates an object specifying one set of analysis choices for the IC Temporal Pattern Discovery method.

Usage

```
createIctpdAnalysis(analysisId = 1, controlPeriodStart = -1080,
  controlPeriodEnd = -361, riskPeriodStart = 1, riskPeriodEnd = 30,
  censor = FALSE, multipleControlPeriods = "110",
  multipleRiskPeriods = "10000", shrinkage = 0.5, icPercentile = 0.025,
  metric = "IC025")
```

4 createIctpdAnalysis

Arguments

analysisId A unique identifier that can later be used to identify the results of this analysis controlPeriodStart

start of the control period - can be set between -99999 and 0, default is -1080.

controlPeriodEnd

end of the control period - can be set between -99999 and 0, default is -361.

riskPeriodStart

start of the risk period - can be set between 0 and 99999, default is 1.

riskPeriodEnd end of the risk period - can be set between 0 and 99999, default is 30.

censor a flag indicating wether the method should censor the observation period at the

end of exposure or not. Available input is 0 or 1 with default = 0.

multipleControlPeriods

Defines the control periods to use where 100 means the control period defined by controlPeriodStart/End, 010 means the period -30 to -1 day before prescription

and 001 means the control period on the day of prescription

multipleRiskPeriods

Defines the risk periods to use 10000 is 1-30 days, 01000 is 1 to 360 days, 00100 is 31 to 90 days, 00010 is 91 to 180 and 00001 is 721 to 1080 days after

prescription default is '10000'

shrinkage used in IRR calculations, required >0 to deal with 0 case counts, but

larger number means more shrinkage. default is 0.5

icPercentile The lower bound of the credibility interval for the IC values (IClow). default is

0.025,

metric Defines wether the output will contain the point estimate or the lower bound.

Available input is 'IC and 'IC025' default is 'IC025'

Examples

```
#Constructing a list of analyses:
analysisList <- NULL
analysis <- createIctpdAnalysis(analysisId = 1, censor = TRUE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis)</pre>
analysis <- createIctpdAnalysis(analysisId = 2, censor = FALSE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis, analysisList)</pre>
#You can store the list:
writeIctpdAnalysisListToFile(analysisList,"ictpdAnalysisList.csv")
#And load it back later:
analysisList <- loadIctpdAnalysisList("ictpdAnalysisList.csv")</pre>
#You can run all analyses across a set of exposure-outcome pairs with one command:
connectionDetails <- createConnectionDetails(dbms="sql server", server="server_ip")</pre>
exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),
                                   exposureConceptId = c(1501700, 1545958, 1551803))
results <- runIctpdAnalyses(connectionDetails,</pre>
                             cdmDatabaseSchema = "cdm_schema.dbo";
                             resultsDatabaseSchema = "results.dbo",
```

getDbIctpdData 5

```
exposureOutcomePairs = exposureOutcomePairs)),
ictpdAnalysisList = ictpdAnalysisList)
```

End(Not run)

getDbIctpdData

Get ICTPD counts from database

Description

This function is used to load the counts needed to compute the ICTPD from a database in OMOP CDM format.

Usage

```
getDbIctpdData(connectionDetails, cdmDatabaseSchema, resultsDatabaseSchema,
  exposureOutcomePairs, exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era", outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "condition_era", drugTypeConceptIdList = c(38000182),
  conditionTypeConceptIdList = c(38000247), controlPeriodStart = -1080,
  controlPeriodEnd = -361, riskPeriodStart = 1, riskPeriodEnd = 30,
  censor = FALSE)
```

Arguments

connectionDetails

An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.

cdmDatabaseSchema

Name of database schema that contains OMOP CDM and vocabulary.

resultsDatabaseSchema

Name of database schema that we can write results to.

exposureOutcomePairs

A data frame with at least two columns:

- "exposureConceptId" containing the drug_concept_ID or cohort_concept_id of the exposure variable
- "outcomeConceptId" containing the condition_concept_ID or cohort_concept_id
 of the outcome variable

exposureDatabaseSchema

The name of the database schema that is the location where the exposure data is available. If exposureTable = DRUG_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

exposureTable

The tablename that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.

outcomeDatabaseSchema

The name of the database schema that is the location where the data used to define the outcome cohorts is available. If exposureTable = CONDITION_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

6 loadIctpdAnalysisList

outcomeTable The tablename that contains the outcome cohorts. If outcome Table <> CONDI-TION OCCURRENCE, then expectation is outcome Table has format of CO-HORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE. drugTypeConceptIdList Which drug_type to use: generally only use 1 value (ex: 30d era). ${\tt conditionTypeConceptIdList}$ Which condition_type to use: generally only use 1 value (ex: 30d era). controlPeriodStart start of the control period - can be set between -99999 and 0, default is -1080. controlPeriodEnd end of the control period - can be set between -99999 and 0, default is -361. riskPeriodStart start of the risk period - can be set between 0 and 99999, default is 1.

riskPeriodEnd end of the risk period - can be set between 0 and 99999, default is 30.

a flag indicating wether the method should censor the observation period at the censor

end of exposure or not. Available input is 0 or 1 with default = 0.

Value

An object of type ictpdData containing counts that can be used in the calculateStatisticsIc function.

Examples

```
## Not run:
connectionDetails <- createConnectionDetails(dbms="sql server", server="server_ip")</pre>
exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),\\
                                    exposureConceptId = c(1501700, 1545958, 1551803))
ictpdData <- getDbIctpdData(connectionDetails,</pre>
                             cdmDatabaseSchema = "cdm_schema.dbo",
                             resultsDatabaseSchema = "results.dbo",
                             exposureOutcomePairs = exposureOutcomePairs)
ictpdResults <- calculateStatisticsIC(ictpdData)</pre>
ictpdResults
## End(Not run)
```

loadIctpdAnalysisList Load a list of Ictpd analysis from file

Description

loadIctpdAnalysisList reads a list of objects of type ictpdAnalysis from a CSV file

Usage

```
loadIctpdAnalysisList(file)
```

Arguments

file the name of the file to be loaded runIctpd 7

Value

A list of objects of type ictpdAnalysis

Examples

```
## Not run:
#Constructing a list of analyses:
analysisList <- NULL
analysis <- createIctpdAnalysis(analysisId = 1, censor = TRUE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis)</pre>
analysis <- createIctpdAnalysis(analysisId = 2, censor = FALSE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis, analysisList)</pre>
#You can store the list:
writeIctpdAnalysisListToFile(analysisList,"ictpdAnalysisList.csv")
#And load it back later:
analysisList <- loadIctpdAnalysisList("ictpdAnalysisList.csv")</pre>
#You can run all analyses across a set of exposure-outcome pairs with one command:
connectionDetails <- createConnectionDetails(dbms="sql server", server="server_ip")</pre>
exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),
                                    exposureConceptId = c(1501700, 1545958, 1551803))
results <- runIctpdAnalyses(connectionDetails,</pre>
                             cdmDatabaseSchema = "cdm_schema.dbo",
                             resultsDatabaseSchema = "results.dbo";
                             exposureOutcomePairs = exposureOutcomePairs)),
                             ictpdAnalysisList = ictpdAnalysisList)
## End(Not run)
```

runIctpd

IC temporal pattern discovery

Description

Ictpd generates population-level estimation from OMOP CDMv4 instance by combining a self-controlled design with a cohort design.

Usage

```
runIctpd(connectionDetails, cdmDatabaseSchema, resultsDatabaseSchema,
  exposureOutcomePairs, exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era", outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "condition_era", drugTypeConceptIdList = c(38000182),
  conditionTypeConceptIdList = c(38000247), storeResultsInDatabase = FALSE,
  createOutputTables = TRUE, outputTablePrefix = "ictpd", analysisId = 1,
  controlPeriodStart = -1080, controlPeriodEnd = -361,
  riskPeriodStart = 1, riskPeriodEnd = 30, censor = FALSE,
  multipleControlPeriods = "110", multipleRiskPeriods = "10000",
  shrinkage = 0.5, icPercentile = 0.025, metric = "IC025")
```

8 runletpd

Arguments

connectionDetails

An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.

cdmDatabaseSchema

Name of database schema that contains OMOP CDM and vocabulary.

resultsDatabaseSchema

Name of database schema that we can write results to.

exposureOutcomePairs

A data frame with at least two columns:

- "exposureConceptId" containing the drug_concept_ID or cohort_concept_id of the exposure variable
- "outcomeConceptId" containing the condition_concept_ID or cohort_concept_id of the outcome variable

exposureDatabaseSchema

The name of the database schema that is the location where the exposure data is available. If exposureTable = DRUG_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

exposureTable The tablename that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.

outcomeDatabaseSchema

The name of the database schema that is the location where the data used to define the outcome cohorts is available. If exposureTable = CONDITION_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

outcomeTable The tablename that contains the outcome cohorts. If outcomeTable <> CONDITION_OCCURRENCE, then expectation is outcomeTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.

 ${\tt drugTypeConceptIdList}$

Which drug type to use: generally only use 1 value (ex: 30d era).

 $\verb|conditionTypeConceptIdList|\\$

Which condition_type to use: generally only use 1 value (ex: 30d era).

storeResultsInDatabase

Should the results be stored in the database?

createOutputTables

Should the output tables be created? If not, they are assumed to exist and data will be appended. If the value is true and the tables exist they will be overwritten.

outputTablePrefix

Prefix used for the result tables in the resultsDatabaseSchema.

analysisId A unique identifier that can later be used to identify the results of this analysis. controlPeriodStart

start of the control period - can be set between -99999 and 0, default is -1080.

controlPeriodEnd

end of the control period - can be set between -99999 and 0, default is -361.

riskPeriodStart

start of the risk period - can be set between 0 and 99999, default is 1.

runIctpdAnalyses 9

riskPeriodEnd end of the risk period - can be set between 0 and 99999, default is 30.

censor a flag indicating wether the method should censor the observation period at the

end of exposure or not. Available input is 0 or 1 with default = 0.

multipleControlPeriods

Defines the control periods to use where 100 means the control period defined by controlPeriodStart/End, 010 means the period -30 to -1 day before prescription

and 001 means the control period on the day of prescription

multipleRiskPeriods

Defines the risk periods to use 10000 is 1-30 days, 01000 is 1 to 360 days, 00100 is 31 to 90 days, 00010 is 91 to 180 and 00001 is 721 to 1080 days after

prescription default is '10000'

shrinkage used in IRR calculations, required >0 to deal with 0 case counts, but

larger number means more shrinkage. default is 0.5

icPercentile The lower bound of the credibility interval for the IC values (IClow). default is

0.025,

metric Defines wether the output will contain the point estimate or the lower bound.

Available input is 'IC and 'IC025' default is 'IC025'

Details

Population-level estimation method that estimates risk by combining a self-controlled and cohort design. This function will call both the getDbIctpdData and calculateStatisticsIc functions.

Value

An object of type ictpdResults containing the results.

References

Noren GN, Bate A, Hopstadius J, Star K, Edwards IR. Temporal Pattern Discovery for Trends and Transient Effect: Its Application to Patient Reocrds. by combining a self-controlled design with a cohort design In: Proceedings of the fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08).ACM Press, New York, pp 963-971

Examples

```
## Not run:
connectionDetails <- createConnectionDetails(dbms="sql server", server="server_ip")
ictpdResults <- runIctpd(connectionDetails, cdmDatabaseSchema, resultsDatabaseSchema, exposureOutcomePairs)
ictpdResults</pre>
```

End(Not run)

runIctpdAnalyses run ICTPD on a list of analysis

Description

This function will run the different ICTPD variations specified in the analysis list on all the exposure-outcome pairs.

10 runIctpdAnalyses

Usage

```
runIctpdAnalyses(connectionDetails, cdmDatabaseSchema, resultsDatabaseSchema,
  exposureOutcomePairs, exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era", outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "condition_era", drugTypeConceptIdList = c(38000182),
  conditionTypeConceptIdList = c(38000247), storeResultsInDatabase = FALSE,
  createOutputTables = TRUE, outputTablePrefix = "ictpd", ictpdAnalysisList)
```

Arguments

connectionDetails

An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.

cdmDatabaseSchema

Name of database schema that contains OMOP CDM and vocabulary.

resultsDatabaseSchema

Name of database schema that we can write results to.

exposureOutcomePairs

A data frame with at least two columns:

- "exposureConceptId" containing the drug_concept_ID or cohort_concept_id of the exposure variable
- "outcomeConceptId" containing the condition_concept_ID or cohort_concept_id
 of the outcome variable

exposureDatabaseSchema

The name of the database schema that is the location where the exposure data is available. If exposureTable = DRUG_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

exposureTable

The tablename that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.

outcomeDatabaseSchema

The name of the database schema that is the location where the data used to define the outcome cohorts is available. If exposureTable = CONDITION_ERA, exposureSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

outcomeTable

The tablename that contains the outcome cohorts. If outcomeTable <> CONDITION_OCCURRENCE, then expectation is outcomeTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.

drugTypeConceptIdList

Which drug_type to use: generally only use 1 value (ex: 30d era).

 ${\tt conditionTypeConceptIdList}$

Which condition_type to use: generally only use 1 value (ex: 30d era).

storeResultsInDatabase

Should the results be stored in the database?

createOutputTables

Should the output tables be created? If not, they are assumed to exist and data will be appended. If the value is true and the tables exist they will be overwritten.

outputTablePrefix

Prefix used for the result tables in the resultsDatabaseSchema.

ictpdAnalysisList

A list of objects of type ictpdAnalysis.

Details

Population-level estimation method that estimates risk by combining a self-controlled and cohort design. This function will call both the getDbIctpdData and calculateStatisticsIc functions.

Value

An object of type ictpdResults containing the results.

References

Noren GN, Bate A, Hopstadius J, Star K, Edwards IR. Temporal Pattern Discovery for Trends and Transient Effect: Its Application to Patient Reocrds. by combining a self-controlled design with a cohort design In: Proceedings of the fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08). ACM Press, New York, pp 963-971

Examples

```
## Not run:
 #Constructing a list of analyses:
 analysisList <- NULL
 analysis <- createIctpdAnalysis(analysisId = 1, censor = TRUE)</pre>
 analysisList <- appendToIctpdAnalysisList(analysis)</pre>
 analysis <- createIctpdAnalysis(analysisId = 2, censor = FALSE)</pre>
 analysisList <- appendToIctpdAnalysisList(analysis, analysisList)</pre>
 #You can store the list:
 writeIctpdAnalysisListToFile(analysisList,"ictpdAnalysisList.csv")
 #And load it back later:
 analysisList <- loadIctpdAnalysisList("ictpdAnalysisList.csv")</pre>
 #You can run all analyses across a set of exposure-outcome pairs with one command:
 connection Details <- \ create Connection Details (dbms="sql server", server="server_ip") \\
 exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),
                                     exposureConceptId = c(1501700, 1545958, 1551803))
 results <- runIctpdAnalyses(connectionDetails,</pre>
                               cdmDatabaseSchema = "cdm_schema.dbo";
                               resultsDatabaseSchema = "results.dbo",
                               exposureOutcomePairs = exposureOutcomePairs)),
                               ictpdAnalysisList = ictpdAnalysisList)
 ## End(Not run)
writeIctpdAnalysisList
                          writeIctpdAnalysisListToFile
```

Description

writeIctpdAnalysisListToFile writes an object of type IctpdAnalysisList to a CSV file

Usage

```
writeIctpdAnalysisList(ictpdAnalysisList, file)
```

Arguments

```
ictpdAnalysisList
the ictpdAnalysisList to be written to file
file the name of the file where the results will be written
```

Examples

```
## Not run:
#Constructing a list of analyses:
analysisList <- NULL
analysis <- createIctpdAnalysis(analysisId = 1, censor = TRUE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis)</pre>
analysis <- createIctpdAnalysis(analysisId = 2, censor = FALSE)</pre>
analysisList <- appendToIctpdAnalysisList(analysis, analysisList)</pre>
#You can store the list:
writeIctpdAnalysisListToFile(analysisList,"ictpdAnalysisList.csv")
#And load it back later:
analysis List <- load Ictpd Analysis List ("ictpd Analysis List.csv")\\
#You can run all analyses across a set of exposure-outcome pairs with one command:
connection Details <- \ create Connection Details (dbms="sql server", server="server_ip") \\
exposureOutcomePairs = data.frame(outcomeConceptId = c(196794, 196794, 312648),
                                    exposureConceptId = c(1501700, 1545958, 1551803))
results <- runIctpdAnalyses(connectionDetails,</pre>
                             cdmDatabaseSchema = "cdm_schema.dbo";
                             resultsDatabaseSchema = "results.dbo";
                             exposureOutcomePairs = exposureOutcomePairs)),
                             ictpdAnalysisList = ictpdAnalysisList)
## End(Not run)
```

Index

```
appendToIctpdAnalysisList, 1
calculateStatisticsIc, 2, 6, 9, 11
createIctpdAnalysis, 3
getDbIctpdData, 3, 5, 9, 11
loadIctpdAnalysisList, 6
runIctpd, 7
runIctpdAnalyses, 9
writeIctpdAnalysisList, 11
```