Package 'MethodEvaluation'

February 17, 2017

Type Package

Title Package for evaluation of estimation methods

Version 0.0.4

Date 2017-02-17

Author Martijn J. Schuemie [aut, cre],

Maintainer Martijn J. Schuemie <schuemie@ohdsi.org>

Description This package contains resources for the evaluation of the performance of methods that aim to estimate the magnitude (relative risk) of the effect of a drug on an outcome. These resources include reference sets for evaluating methods on real data, as well as functions for inserting simulated effects in real data based on negative control drug-outcome pairs. Further included are functions for the computation of the minimum detectable relative risks and functions for computing performance statistics such as predictive accuracy, error and bias.

License Apache License 2.0

Depends R (>= 3.2.0),
DatabaseConnector (>= 1.3.0),
FeatureExtraction,
Cyclops (>= 1.2.3)

Imports ff,
ffbase (>= 0.12.1),
RJDBC,
SqlRender (>= 1.1.2),
pROC,
ggplot2,
OhdsiRTools

Suggests testthat

RoxygenNote 5.0.1

R topics documented:

`computeAuc` *Compute the area under the ROC curve*

Description

Compute the area under the ROC curve

Usage

```
computeAuc(methodResults, referenceSet, confidenceIntervals = TRUE)
```

`computeAucs` *Compute the AUCs for various injected signal sizes*

Description

Compute the AUCs for various injected signal sizes

Usage

```
computeAucs(logRr, trueLogRr)
```

Arguments

`logRr`	A vector containing the log of the relative risk as estimated by a method.
`trueLogRr`	A vector containing the injected log(relative risk) for each estimate.

Value

A data frame with per injected signal size the AUC and the 95 percent confidence interval of the AUC.

computeCoverage *Compute the coverage*

Description

Compute the coverage

Usage

```
computeCoverage(logRr, seLogRr, trueLogRr, region = 0.95)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale.
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025).
trueLogRr	A vector of the true effect sizes.
region	Size of the confidence interval. Default is .95 (95 percent).

Details

Compute the fractions of estimates where the true effect size is below, above or within the confidence interval, for one or more true effect sizes.

computeMdrr *Compute minimal detectable relative risk (MDRR)*

Description

computeMdrr computes the minimal detectable relative risk (MDRR) for drug-outcome pairs.

Usage

```
computeMdrr(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema, exposureOutcomePairs,
  exposureDatabaseSchema = cdmDatabaseSchema, exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "condition_era",
  cdmVersion = "5")
```

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.

`oracleTempSchema`
: For Oracle only: the name of the database schema where you want all temporary tables to be managed. Requires create/insert permissions to this database.

`exposureOutcomePairs`
: A data frame with at least two columns:
  - "exposureConceptId" containing the drug_concept_ID or cohort_definition_id of the exposure variable
  - "outcomeConceptId" containing the condition_concept_ID or cohort_definition_id of the outcome variable

`exposureDatabaseSchema`
: The name of the database schema that is the location where the exposure data used to define the exposure cohorts is available. If exposureTable = DRUG_ERA, exposureDatabaseSchema 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, exposureDatabaseSchema 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.

`cdmVersion`
: Define the OMOP CDM version used: currently support "4" and "5".

Details

Computes the MDRR using simple power-calculations using person-level statistics stratified by age and gender.

Value

A data frame containing the MDRRs for the given exposure-outcome pairs.

Examples

```
## Not run:
connectionDetails <- createConnectionDetails(dbms = "sql server",
                                             server = "RNDUSRDHIT07.jnj.com")
exposureOutcomePairs <- data.frame(exposureConceptId = c(767410, 1314924, 907879),
                                   outcomeConceptId = c(444382, 79106, 138825))
mdrrs <- computeMdrr(connectionDetails,
                     "cdm_truven_mdcr",
                     exposureOutcomePairs,
                     outcomeTable = "condition_era")

## End(Not run)
```

computeMetrics *Compute the AUC, coverage, MSE, and type 1 and 2 error*

Description

Compute the AUC, coverage, MSE, and type 1 and 2 error

Usage

```
computeMetrics(logRr, seLogRr, trueLogRr)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025)
trueLogRr	A vector of the true effect sizes

Details

Compute the AUC, coverage, MSE, and type 1 and 2 error.

computeMse *Compute the mean squared error*

Description

Compute the mean squared error

Usage

```
computeMse(logRr, trueLogRr)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale.
trueLogRr	A vector of the true effect sizes.

computeType1And2Error *Compute type 1 and 2 error*

Description

Compute type 1 and 2 error

Usage

```
computeType1And2Error(logRr, seLogRr, trueLogRr, alpha = 0.05)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale.
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025).
trueLogRr	A vector of the true effect sizes.
alpha	The alpha (expected type I error).

createOutcomeCohorts *Create outcomes of interest*

Description

Create outcomes of interest

Usage

```
createOutcomeCohorts(connectionDetails, cdmDatabaseSchema,
  createNewCohortTable = FALSE, cohortDatabaseSchema = cdmDatabaseSchema,
  cohortTable = "cohort", referenceSet = "omopReferenceSet")
```

Arguments

connectionDetails	An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cdmDatabaseSchema	A database schema containing health care data in the OMOP Commond Data Model. Note that for SQL Server, botth the database and schema should be specified, e.g. 'cdm_schema.dbo'
createNewCohortTable	Should a new cohort table be created, or should the outcomes be inserted in a existing table?
cohortDatabaseSchema	The database schema where the target table is located. Note that for SQL Server, botth the database and schema should be specified, e.g. 'cdm_schema.dbo'
cohortTable	The name of the table where the outcomes will be stored.
referenceSet	The name of the reference set for which outcomes need to be created.

Details

This function will create the outcomes of interest referenced in the various reference sets. The outcomes of interest are derives using information like diagnoses, procedures, and drug prescriptions. The outcomes are stored in a table on the database server.

`euadrReferenceSet` *The EU-ADR reference set*

Description

A reference set of 43 drug-outcome pairs where we believe the drug causes the outcome (positive controls) and 50 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 10 health outcomes of interest. Note that originally, there was an additional positive control (Nimesulide and acute liver injury), but Nimesulide is not in RxNorm, and is not available in many countries.

Usage

```
data(euadrReferenceSet)
```

Format

A data frame with 399 rows and 10 variables:

exposureConceptId Concept ID identifying the exposure

exposureConceptName Name of the exposure

outcomeConceptId Concept ID identifying the outcome

outcomeConceptName Name of the outcome

groundTruth 0 = negative control, 1 = positive control

indicationConceptId Concept Id identifying the (primary) indication of the drug. To be used when one wants to nest the analysis within the indication

indicationConceptName Name of the indication

comparatorDrugConceptId Concept ID identifying a comparator drug that can be used as a counterfactual

comparatorDrugConceptName Name of the comparator drug

comparatorType How the comparator was selected

References

Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, Fourrier-Reglat A, Molokhia M, Patadia V, van der Lei J, Sturkenboom M, Trifiro G. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. Drug Safety 36(1):13-23, 2013

filterOnMdrr *Filter data based on MDRR*

Description

Filters a dataset to those exposure-outcome pairs with sufficient power.

Usage

```
filterOnMdrr(data, mdrr, threshold = 1.25)
```

Arguments

data	A data frame with at least two columns: • "exposureConceptId" containing the drug_concept_ID or cohort_definition_id of the exposure variable • "outcomeConceptId" containing the condition_concept_ID or cohort_definition_id of the outcome variable
mdrr	A data frame as generated by the computeMdrr function.
threshold	The required minimum detectable relative risk.

Value

A subset of the data object.

injectSignals *Inject signals in database*

Description

Inject signals in database

Usage

```
injectSignals(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema,
  exposureDatabaseSchema = cdmDatabaseSchema, exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "cohort",
  outputDatabaseSchema = outcomeDatabaseSchema, outputTable = outcomeTable,
  createOutputTable = FALSE, exposureOutcomePairs, modelType = "poisson",
  buildOutcomeModel = TRUE, buildModelPerExposure = FALSE,
  minOutcomeCountForModel = 100, minOutcomeCountForInjection = 25,
  covariateSettings = FeatureExtraction::createCovariateSettings(useCovariateDemographics
  = TRUE, useCovariateDemographicsGender = TRUE, useCovariateDemographicsRace =
  TRUE, useCovariateDemographicsEthnicity = TRUE, useCovariateDemographicsAge =
  TRUE, useCovariateDemographicsYear = TRUE, useCovariateDemographicsMonth =
  TRUE, useCovariateConditionOccurrence = TRUE,
  useCovariateConditionOccurrence365d = TRUE, useCovariateConditionOccurrence30d
  = TRUE, useCovariateConditionOccurrenceInpt180d = TRUE,
```

```
  useCovariateConditionEra = TRUE,      useCovariateConditionEraEver = TRUE,
  useCovariateConditionEraOverlap = TRUE, useCovariateConditionGroup = TRUE,
  useCovariateDrugExposure = TRUE, useCovariateDrugExposure365d = TRUE,
  useCovariateDrugExposure30d = TRUE, useCovariateDrugEra = TRUE,
  useCovariateDrugEra365d = TRUE, useCovariateDrugEra30d = TRUE,
  useCovariateDrugEraEver = TRUE, useCovariateDrugEraOverlap = TRUE,
  useCovariateDrugGroup = TRUE, useCovariateProcedureOccurrence = TRUE,
 useCovariateProcedureOccurrence365d = TRUE, useCovariateProcedureOccurrence30d
  = TRUE,      useCovariateProcedureGroup = TRUE, useCovariateObservation =
  TRUE, useCovariateObservation365d = TRUE, useCovariateObservation30d = TRUE,
  useCovariateObservationCount365d = TRUE, useCovariateMeasurement365d = TRUE,
  useCovariateMeasurement30d = TRUE, useCovariateMeasurementCount365d = TRUE,
  useCovariateMeasurementBelow = TRUE, useCovariateMeasurementAbove = TRUE,
  useCovariateConceptCounts = TRUE, useCovariateRiskScores = TRUE,
  useCovariateRiskScoresCharlson = TRUE, useCovariateRiskScoresDCSI = TRUE,
  useCovariateRiskScoresCHADS2 = TRUE, useCovariateRiskScoresCHADS2VASc = TRUE,
  useCovariateInteractionYear = FALSE, useCovariateInteractionMonth = FALSE,
  excludedCovariateConceptIds = c(), deleteCovariatesSmallCount = 100),
  prior = createPrior("laplace", exclude = 0, useCrossValidation = TRUE),
  control = createControl(cvType = "auto", startingVariance = 0.1, noiseLevel
  = "quiet", threads = 10), firstExposureOnly = FALSE, washoutPeriod = 183,
  riskWindowStart = 0, riskWindowEnd = 0, addExposureDaysToEnd = TRUE,
  firstOutcomeOnly = FALSE, removePeopleWithPriorOutcomes = FALSE,
  maxSubjectsForModel = 1e+05, effectSizes = c(1, 1.25, 1.5, 2, 4),
  precision = 0.01, outputIdOffset = 1000,
  workFolder = "./SignalInjectionTemp", cdmVersion = "4",
  modelThreads = 1, generationThreads = 1)
```

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.

`oracleTempSchema`
: For Oracle only: the name of the database schema where you want all temporary tables to be managed. Requires create/insert permissions to this database.

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

`exposureTable`
: The table name that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: cohort_concept_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, exposureDatabaseSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

`outcomeTable` The table name that contains the outcome cohorts. When the table name is not CONDITION_ERA This table is expected to have the same format as the COHORT table: SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE, COHORT_CONCEPT_ID (CDM v4) or COHORT_DEFINITION_ID (CDM v5 and higher).

`outputDatabaseSchema`
The name of the database schema that is the location of the tables containing the new outcomesRequires write permissions to this database.

`outputTable` The name of the table names that will contain the generated outcome cohorts.

`createOutputTable`
Should the output table be created prior to inserting the outcomes? If TRUE and the tables already exists, it will first be deleted. If FALSE, the table is assumed to exist and the outcomes will be inserted. Any existing outcomes with the same IDs will first be deleted.

`exposureOutcomePairs`
A data frame with at least two columns:

- "exposureId" containing the drug_concept_ID or cohort_concept_id of the exposure variable
- "outcomeId" containing the condition_concept_ID or cohort_concept_id of the outcome variable

`modelType` Can be either "poisson" or "survival"

`buildOutcomeModel`
Should an outcome model be created to predict outcomes. New outcomes will be inserted based on the predicted probabilities according to this model, and this will help preserve the observed confounding when injecting signals.

`buildModelPerExposure`
If TRUE, an outcome model will be created for each exposure ID. IF false, outcome models will be created across all exposures.

`minOutcomeCountForModel`
Minimum number of outcome events required to build a model.

`minOutcomeCountForInjection`
Minimum number of outcome events required to inject a signal.

`covariateSettings`
An object of type `covariateSettings` as created using the `createCovariateSettings` function in the `FeatureExtraction` package.

`prior` The prior used to fit the outcome model. See `createPrior` for details.

`control` The control object used to control the cross-validation used to determine the hyperparameters of the prior (if applicable). See `createControl` for details.

`firstExposureOnly`
Should signals be injected only for the first exposure? (ie. assuming an acute effect)

`washoutPeriod` Number of days at the start of observation for which no signals will be injected, but will be used to determine whether exposure or outcome is the first one, and for extracting covariates to build the outcome model.

`riskWindowStart`
The start of the risk window relative to the start of the exposure (in days). When 0, risk is assumed to start on the first day of exposure.

`riskWindowEnd` The end of the risk window relative to the start of the exposure. Note that typically the length of exposure is added to this number (when the `addExposureDaysToEnd` parameter is set to TRUE).

`addExposureDaysToEnd`
Should length of exposure be added to the risk window?

`firstOutcomeOnly`
Should only the first outcome per person be considered when modeling the outcome?

`removePeopleWithPriorOutcomes`
Remove people with prior outcomes?

`maxSubjectsForModel`
Maximum number of people used to fit an outcome model.

`effectSizes` A numeric vector of effect sizes that should be inserted.

`precision` The allowed ratio between target and injected signal size.

`outputIdOffset` What should be the first new outcome ID that is to be created?

`workFolder` Path to a folder where intermediate data will be stored.

`cdmVersion` Define the OMOP CDM version used: currently support "4" and "5".

`modelThreads` Number of parallel threads to use when fitting outcome models.

`generationThreads`
Number of parallel threads to use when generating outcomes.

Details

This function will insert additional outcomes for a given set of drug-outcome pairs. It is assumed that these drug-outcome pairs represent negative controls, so the true relative risk before inserting any outcomes should be 1. There are two models for inserting the outcomes during the specified risk window of the drug: a Poisson model assuming multiple outcomes could occurr during a single exposure, and a survival model considering only one outcome per exposure. For each

Value

A data.frame listing all the drug-pairs in combination with requested effect sizes and the real inserted effect size (might be different from the requested effect size because of sampling error).

`MethodEvaluation` *MethodEvaluation*

Description

MethodEvaluation

`omopReferenceSet` *The OMOP reference set A reference set of 165 drug-outcome pairs where we believe the drug causes the outcome (positive controls) and 234 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 4 health outcomes of interest: acute liver injury, acute kidney injury, acute myocardial infarction, and GI bleeding.*

Description

The OMOP reference set A reference set of 165 drug-outcome pairs where we believe the drug causes the outcome (positive controls) and 234 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 4 health outcomes of interest: acute liver injury, acute kidney injury, acute myocardial infarction, and GI bleeding.

Usage

```
data(omopReferenceSet)
```

Format

A data frame with 399 rows and 10 variables:

exposureConceptId Concept ID identifying the exposure

exposureConceptName Name of the exposure

outcomeConceptId Concept ID identifying the outcome

outcomeConceptName Name of the outcome

groundTruth 0 = negative control, 1 = positive control

indicationConceptId Concept Id identifying the (primary) indication of the drug. To be used when one wants to nest the analysis within the indication

indicationConceptName Name of the indication

comparatorDrugConceptId Concept ID identifying a comparator drug that can be used as a counterfactual

comparatorDrugConceptName Name of the comparator drug

comparatorType How the comparator was selected

References

Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Safety 36 Suppl 1:S33-47, 2013

plotCoverageInjectedSignals *Plot the coverage*

Description

Plot the coverage

Usage

```
plotCoverageInjectedSignals(logRr, seLogRr, trueLogRr, region = 0.95,
  fileName = NULL)
```

Arguments

logRr	A numeric vector of effect estimates on the log scale
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025)
trueLogRr	A vector of the true effect sizes
region	Size of the confidence interval. Default is .95 (95 percent).
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggsave in the ggplot2 package for supported file formats.

Details

Plot the fractions of estimates where the true effect size is below, above or within the confidence interval, for one or more true effect sizes.

plotRocsInjectedSignals *Plot the ROC curves for various injected signal sizes*

Description

Plot the ROC curves for various injected signal sizes

Usage

```
plotRocsInjectedSignals(logRr, trueLogRr, showAucs, fileName = NULL)
```

Arguments

logRr	A vector containing the log of the relative risk as estimated by a method.
trueLogRr	A vector containing the injected log(relative risk) for each estimate.
showAucs	Should the AUCs be shown in the plot?
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggsave in the ggplot2 package for supported file formats.

Value

A Ggplot object. Use the `ggsave` function to save to file.

Index