mitoticFigureCounts

October 13, 2019

Type Package

Title Mitotic figure counts study data, functions, and markdown files (Tabata2019_Diagn-Pathol_v14p65)

Version 1.0

Date 2019-10-12

Author Brandon D Gallas, Kazuhiro Tabata, Qi Gong

Maintainer Dr. Brandon D. Gallas <brandon.gallas@fda.hhs.gov>

Description This package contains the data, functions, and markdown files for a study comparing mitotic figure counting performances based on whole slide images (WSI images) from four scanners and a microscope. The details about this study are described in the following paper:
Tabata, K., N. Uraoka, J. Benhaminda, M. G. Hanna, S.J. Sirintrapun, B. D. Gallas, Q. Gong, R. G. Aly, K. Emoto and K. M. J. D. P. Matsuda (2019). ``Validation of mitotic cell quantification via microscopy and multiple whole-slide scanners,'' Diagn Pathol, 14(1): 65.

License CC0

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

Depends R (>= 2.10)

R topics documented:

aucMRMCcluster *aucMRMCcluster*

Description

The results of a multi-reader multi-case (MRMC) analysis of the auc for each scanner. The MRMC analysis is accomplished by the OR method (Obuchowski and Rockette: Obuchowski1995_CommunStat-Simulat_v24p285). The function used is `mrmcAnalysisOR`. Since the data is binary, auc is the average of sensitivity and specificity or half of (Youden's index + 1). Sensitivity is defined as the number of MFs detected by an observer divided by the number of true MFs. Specificity is defined as one minus the false-positive fraction, where the false-positive fraction is the number of false MFs that were positively marked, divided by the total number of false MFs. Furthermore, we account for the fact that there are multiple observations per case (multiple ROIs per WSI, clustered data: Obuchowski1997_Biometrics_v53p567) when calculating the reader by modality covariances that are used in the OR method. The function used is `doAUCcluster`

Usage

```
aucMRMCcluster
```

Format

An object of class `list` of length 4.

Details

The results of this script are a list of MRMC AUC analysis results for 4 scanners.

- Scanner.A,
- Scanner.B,
- Scanner.C,
- Scanner.D

Each scanner results is itself a list contains the MRMC analysis results for the null hypothesis that the scanner accuracy is the same as the microscope accuracy:

- `F_stat`, F-statistic
- `df.H`, Degrees of freedom
- `p`, p-value
- `theta.i`, [2] Accuracy of the microscope and scanner
- `df.sgl`, [2] Degrees of freedom for the microscope and the scanner
- `se.i`, [2] Standard error of the microscope and scanner accuracy
- `se.dif`, Standard erro of the microscope and scanner accuracy difference
- `covOR`, [12] Obuchowski and Rockette covariances: varR, varTR, cov1, cov2, cov3, varE, varR.1, varR.2, cov2.1, cov2.2, varE.1, varE.2
- `theta.hat`, [2x5] Accuracy of each reader for the microscope (row 1) and accuracy of each reader for the scanner (row 2)
- `cov.hat`, [10x10] Reader by modality (microscope and scanner) covariance matrix

- botCI, [2] Bottom 95
- topCI, [2] Top 95

#' The analysis results are used to produce:

- Table 4: Accuracy for all readers and observation methods
- Figure 3: Accuracy (average of sensitivity and specificity) for each viewing mode averaged over all the readers with 95 The asterisks indicate that the difference in accuracy of the viewing mode compared to that of microscopy is statistically significant. All analyses account for the correlations and variability from the readers reading the same ROIs.

---

convertDF *convertDF*

Description

This function converts a "matrixWithTruth" data frame to a "listWithTruth" data frame. The "matrix" part of the input data frame is understood to contain scores from readers: One column for each reader. Column names are the reader names. The reader names (column names) need to be specified in the "readersVector" input parameter.

Usage

```
convertDF(inDF, inDFtype, outDFtype, readers, nameTruth)
```

Arguments

inDF	Counting data frame: dfClassify: dfClassify20180627
inDFtype	Type of input data frame
outDFtype	Type of output data frame
readers	Vector of reader IDs
nameTruth	Column name of truth

Value

Output data frame with 11 column targetID, cell.mark, wsiName, roiID, cellID.mskcc20171103, xCell, yCell, truth, modalityID, readerID, score

Examples

```
dfClassify <- mitoticFigureCounts::dfClassify20180627
readersVector <- names(dfClassify)[8:12]
nameTruth <- names(dfClassify)[13]
df.convert <- convertDF(dfClassify, "matrixWithTruth", "listWithTruth", readersVector, nameTruth)
head(dfClassify)
head(df.convert)
```

dfClassify20180627 *dfClassify20180627*

Description

A single data frame of the study data. Each row corresponds to a candidate mitotic figure and modality (155 candidates * 5 modalities = 775 rows)

Usage

```
dfClassify20180627
```

Format

An object of class `data.frame` with 775 rows and 14 columns.

Details

A data frame with 14 columns:

- `targetID` [Factor] Target cell ID = (candidate mitotic figure ID)
- `cell.mark` [Factor] ROI ID, "dash", Cell ID in the ROI
- `wsiName` [Factor] WSI file name
- `roiID` [Factor] ROI ID
- `cellID.mskcc20171103` [Factor] Descriptive target cell ID, including ROI ID, WSI file name, x position of target cell, y position of target cell
- `xCell` [Factor] x position of target cell
- `yCell` [Factor] y position of target cell
- `observer.1` [num] Observer1 modality-specific decision on the candidate MF
- `observer.2` [num] Observer2 modality-specific decision on the candidate MF
- `observer.3` [num] Observer3 modality-specific decision on the candidate MF
- `observer.4` [num] Observer4 modality-specific decision on the candidate MF
- `observer.5` [num] Observer5 modality-specific decision on the candidate MF
- `truth` [num] Truth panel decision on the candidate MF
- `modalityID` [Factor] modality ID (scanner or microscope)

Some of this data is missing location information.

dfCountROI20180627 *dfCountROI20180627*

Description

A single data frame of the mitotic figure counts per ROI and modality (40 ROIs x 5 modalities = 200 rows). There is a column for each observer and the truth.

Usage

```
dfCountROI20180627
```

Format

An object of class `data.frame` with 200 rows and 9 columns.

Details

A data frame with 9 columns:

- `wsiName` [Factor] WSI file name
- `roiID` [Factor] ROI ID
- `modalityID` [Factor] modality, scanner ID
- `observer.1` [num] Observer1 modality-specific count for the ROI
- `observer.2` [num] Observer2 modality-specific count for the ROI
- `observer.3` [num] Observer3 modality-specific count for the ROI
- `observer.4` [num] Observer4 modality-specific count for the ROI
- `observer.5` [num] Observer5 modality-specific count for the ROI
- `truth` [num] Truth panel count for the ROI

dfCountWSI20180627 *dfCountWSI20180627*

Description

A single data frame of the mitotic figure counts per WSI and modality (4 WSIs x 5 modalities = 20 rows). There is a column for each observer and the truth.

Usage

```
dfCountWSI20180627
```

Format

An object of class `data.frame` with 20 rows and 8 columns.

Details

A data frame with 8 columns:

- `wsiName` [Factor] WSI file name
- `modalityID` [Factor] modality, scanner ID
- `observer.1` [num] Observer1 modality-specific count for the WSI
- `observer.2` [num] Observer2 modality-specific count for the WSI
- `observer.3` [num] Observer3 modality-specific count for the WSI
- `observer.4` [num] Observer4 modality-specific count for the WSI
- `observer.5` [num] Observer5 modality-specific count for the WSI
- `truth` [num] Truth panel count for the WSI

`doAUCcluster` *doAUCcluster*

Description

Do AUC analysis of clustered data.

This function is based on the analysis described in: Obuchowski NA. "Nonparametric analysis of clustered ROC curve data." Biometrics. 1997: 567-578.

This function is an adaptation of a function downloaded from the Cleveland Clinic Lerner Research Institute Department of Quantitative Health Sciences Software web page.

FILE: https://www.lerner.ccf.org/qhs/software/lib/funcs_clusteredROC.R

WEBPAGE: https://www.lerner.ccf.org/qhs/software/roc_analysis.php

Usage

```
doAUCcluster(predictor1, predictor2 = NULL, response, clusterID,
  alpha = 0.05, level = NULL, print.all = F)
```

Arguments

`predictor1`	a vector containing the predictor for ROC curve 1
`predictor2`	a vector containing the predictor for ROC curve 2
`response`	a vector containing the response for both ROC curves
`clusterID`	a vector containing IDs for the clusters
`alpha`	the type I error rate
`level`	can be used to specify the response level considered positive (if omitted, the second level of the response is selected)
`print.all`	if TRUE, intermediate estimates are printed

Details

iMRMC users shared the links during a discussion with questions about how to analyze MRMC data that was clustered. https://github.com/DIDSR/iMRMC/issues/147 There is a short pdf tutorial a https://www.lerner.ccf.org/qhs/software/lib/clusteredROC_help.pdf. It exists in the inst/extra/docs folder of the repository. It exists in the extra/docs folder of the installed package.

Value

list auc, auc.se, ci.for.auc

`mrmcAnalysisOR` *MRMC analysis by Obuchowski and Rockette (Obuchowski1995_Commun-Stat-Simulat_v24p285)*

Description

OR's method to estimate the MRMC variance for a given theta. This function requires an nModalities by nReaders covariance matrix to estimate covOR = (cov1, cov2, cov3), as described by Hillis (2014, SIM, Section 2.2)

Usage

```
mrmcAnalysisOR(theta.hat, cov.hat, sgl = "corres")
```

Arguments

`theta.hat`	[2, nReaders]: Performance estimates for two modalities by all readers
`cov.hat`	[2 * nReaders, 2 * nReaders]: Covariance matrix between each reader x modality
`sgl`	ch: Flat indicating whether or not to do single modality analysis only using single modality data

Value

list

- `F` : F statistic
- `df.H` : degrees of freedom difference of concordance
- `p` : P value
- `se.dif` : difference of variance
- `theta.i` [2] : reader-averaged HarrellsC form modality A & B
- `df.sgl` [2] : degrees of freedom for modality A & B
- `se.i` [2] : Variance form modality A & B
- `covOR` [6] : Components of variance of the Obuchowski and Rockette method
  - cov1, cov2(pooled over modalities), cov3,
  - varC, cov2(modalityA only), cov2(modalityB only)

Index