|  |  |
| --- | --- |
| RJafroc-package {RJafroc} | R Documentation |

R implementation of jackknife alternative free-response receiver operating characteristic analysis

**Description**

This package implements software for the analysis of data acquired in observer performance studies conducted using ROC, FROC or ROI data collection paradigms. It is an R implementation of traditional JAFROC analysis ([www.devchakraborty.com](http://127.0.0.1:21213/help/library/RJafroc/html/www.devchakraborty.com)) with enhancements including allowing choice between DBM or OR significance testing, with Hillis improvements, and choice between several figures of merit, plotting empirical or search model predicted operating characteristics, 3 sample size estimation tools and generating a formatted output.

**Details**

|  |  |
| --- | --- |
| Package: | RJafroc |
| Type: | Package |
| Version: | 1.0 |
| Date: | 2014-12-09 |
| License: | GPL-3 |
| URL: | [www.devchakraborty.com](http://127.0.0.1:21213/help/library/RJafroc/html/www.devchakraborty.com) |
|  |  |

**Abbreviations and definitions**

* ROC: receiver operating characteristic
* FROC: free-response ROC
* AFROC: alternative FROC
* JAFROC: jackknife AFROC
* ROI: region-of-interest
* FOM: figure of merit or quantitative measure of performance
* FP: false positive
* TP: true positive
* FPF: number of FPs divided by number of non-diseased cases
* TPF: number of TPs divided by number of diseased cases
* ROC curve: plot of TPF vs. FPF
* AUC: trapezoidal area under the ROC curve as estimated by the Wilcoxon statistic
* NL: non-lesion localization, of which FP is a special case, i.e., a mark that does not correctly locate any existing localized lesion(s)
* LL: lesion localization, of which TP is a special case, i.e., a mark that correctly locates an existing localized lesion
* LLF: number of LLs divided by the total number of lesions
* NLF: number of NLs divided by the total number of cases
* FROC curve: plot of LLF vs. NLF
* AFROC curve: plot of LLF vs. FPF, where FPF is inferred using highest rating of NL marks on non-diseased cases only
* AFROC1 curve: plot of LLF vs. FPF1, where FPF1 is inferred using highest rating of NL marks on all cases
* JAFROC FOM: trapezoidal area under AFROC curve
* JAFROC1 FOM: trapezoidal area under AFROC1 curve
* CI: confidence interval
* SP: specificity
* SE: sensitivity
* I: total number of modalities, indexed by *i*; *I* must be at least 2 to perform the null hypothesis testing
* J: total number of readers, indexed by *j*
* K1: total number of non-diseased cases, indexed by *k1*
* K2: total number of diseased cases, indexed by *k2*
* K: total number of cases, *K = K1 + K2*, indexed by *k*
* nNL: maximum number of NL marks per case in dataset, indexed by *n*
* sLL: maximum number of LL marks per case in dataset , indexed by *s*

**Dataset**

Dataset, an R object, can be created by the users or read from an external data file with specified format that is described in next section. **Note** that the word "dataset" used in this whole package represents the R object with following structure.

***Data structure***

The structure of the dataset used in analysis functions is an R list containing 9 elements as follows. **Note** that -Inf is assigned to any missing/unavailabe element.

* **NL**: a [I, J, K, nNL] floating-point array that contains the ratings of all NL marks. For ROC datasets FP ratings are assigned to **NL** with *n = 1*.
* **LL**: a [I, J, K2, sLL] floating-point array that contains the ratings of all LL marks. For ROC datasets TP ratings are assigned to **LL** with *s = 1*.
* **lesionNum**: a [K2] integer array, with elements indicating the number of lesions in each diseased case.
* **lesionID**: a [K2, sLL] integer array. **Note** that ratings of lesions in **LL** must appear in the same sequence as **lesionID** for that case. For example, if the **lesionID** field for the first diseased case is c(4, 2, 3, 1), i.e., there are 4 lesion on this case labeled 4,3 2 and 1, the ratings in **LL** for this case must appear in the same sequence, with the first rating corresponding to the lesion labeled 4, the second corresponding to the lesion labeled 2, etc.
* **lesionWeight**: a floating point [K2, sLL] array , representing the relative importance of detecting each lesion. For each case, the weights must sum to unity. If zero is assigned to all elements of this array, then the software assigns equal weighting.
* **maxNL**: the maximum number of NL marks per case in dataset.
* **dataType**: a string variable: "ROC", "ROI" or "FROC".
* **modalityID**: a string vector of length *I*, which labels the modalities in the dataset.
* **readerID**: a string vector of length *J*, which contains the ID of each reader. **Note** that the order of elements in **modalityID** and **readerID** must match that in **NL** and **LL**. For example, NL[1, 2, , ] indicates the ratings of the reader with the second ID in **readerID** using the modality with the first ID in **modalityID**.

***Data file format***

The package reads JAFROC, OR-DBM MRMC (ROC data only) and iMRMC (ROC data only) data files. The data can be imported by using the function [ReadDataFile](http://127.0.0.1:21213/help/library/RJafroc/help/ReadDataFile).

**JAFROC data file format**

The JAFROC data file is an Excel file containing three worksheets (\*.xls and \*.xlsx are supported): (1) the Truth worksheet, (2) the TP or lesion localization worksheet and (3) the FP or non-lesion localization worksheet. Except for the Truth worksheet, where each case must occur at least once, the number of rows in the other worksheets is variable.

1. Truth worksheet consists of
   * CaseID, an integer field uniquely labeling the cases (images). It must occur at least once for each case, and since a case may have multiple lesions, it can occur multiple times, once for each lesion.
   * LesionID, an integer field uniquely labeling the lesions in each case. This field is zero for non-diseased cases.
   * Weight, a floating-point field, which is the relative importance of detecting each lesion. This field is zero for non-diseased cases and for equally weighted lesions; otherwise the weights must sum to unity for each case. Unless a weighted figure of merit is selected, this field is irrelevant.
2. TP worksheet consists of
   * ReaderID, a string field uniquely labeling the readers (radiologists).
   * ModalityID, a string field uniquely labeling the modalities.
   * CaseID, see Truth worksheet. A non-diseased case in this field will generate an error.
   * LesionID, see Truth worksheet. An entry in this field that does not appear in the Truth worksheet will generate an error. It is the user's responsibility to ensure that the entries in the Truth and TP worksheets correspond to the same physical lesions.
   * TP\_Rating, a positive floating-point field denoting the rating assigned to a particular lesion-localization mark, with higher numbers represent greater confidence that the location is actually a lesion.
3. FP worksheet consists of
   * ReaderID, see TP worksheet.
   * ModalityID, see TP worksheet.
   * CaseID, see TP worksheet.
   * FP\_Rating, a positive floating-point field denoting the rating assigned to a particular lesion-localization mark, with higher numbers represent greater confidence that the location is actually a lesion.

**OR-DBM MRMC data file format / LABMRMC format**

* **Input format for OR-DBM MRMC.** This format is described in the Medical Image Perception Laboratory website, currently <http://perception.radiology.uiowa.edu/>.
* **LABMRMC data format.** The data file includes following parts. The file must be saved as plain text file with \*.lrc extension. All items in the file are separated by one or more blank spaces.
  1. The first line is a free text description of the file.
  2. The second line is the name or ID of the first reader.
  3. The third line has the names or IDs of all the modalities. Each name or ID must be enclosed by double quotes(" ").
  4. The fourth line must have the letter (l or s) or word (large or small) for each modality. The letter or word indicates that smaller or larger ratings represent stronger confidence of presence of disease.
  5. The following lines contain the ratings in all modalities, separated by spaces or tabs, of the non-diseased cases, one case per line. The cases must appear in the same order for all readers. Missing value is not allowed.
  6. After the last non-diseased case insert a line containing the asterisk (\*) symbol.
  7. Repeat steps 5 and 6 for the diseased cases.
  8. Repeat steps 2, 5, 6 and 7 for the remaining readers.
  9. The last line of the data file must be a pound symbol (#).

**iMRMC data format**

This format is described in the iMRMC website, currently <https://code.google.com/p/imrmc/>.

**Functions**

|  |  |
| --- | --- |
| [CalculateFOM](http://127.0.0.1:21213/help/library/RJafroc/help/CalculateFOM) | Calculate the figure of merit for each reader using each modality. |
| [DBMHAnalysis](http://127.0.0.1:21213/help/library/RJafroc/help/DBMHAnalysis) | Performs Dorfman-Berbaum-Metz analysis with Hillis improvements for the specified dataset. |
| [OperatingCharacteristics](http://127.0.0.1:21213/help/library/RJafroc/help/OperatingCharacteristics) | Visualize the predicted ROC, AFROC, FROC and pdf (probability density function) curves for given search model parameters. |
| [ORHAnalysis](http://127.0.0.1:21213/help/library/RJafroc/help/ORHAnalysis) | Performs Obuchowski-Rockette analysis with Hillis improvements for the specified dataset. |
| [OutputReport](http://127.0.0.1:21213/help/library/RJafroc/help/OutputReport) | Save the results of the analysis to a text file. |
| [PlotEmpiricalCurve](http://127.0.0.1:21213/help/library/RJafroc/help/PlotEmpiricalCurve) | Plot empirical curves for specified modalities and readers in the dataset. |
| [PlotParametricCurve](http://127.0.0.1:21213/help/library/RJafroc/help/PlotParametricCurve) | Plot parametric curves for specified modalities and readers in the dataset. |
| [PowerGivenJK](http://127.0.0.1:21213/help/library/RJafroc/help/PowerGivenJK) | Calculate the statistical power with the given number of readers, number of cases and DBM or OR variances components. |
| [PowerTable](http://127.0.0.1:21213/help/library/RJafroc/help/PowerTable) | Calculate required sample size for the specified dataset with given significance level, effect size and desired power. |
| [ReadDataFile](http://127.0.0.1:21213/help/library/RJafroc/help/ReadDataFile) | Read the dataset that will be analysis from data file. |
| [SampleSizeGivenJ](http://127.0.0.1:21213/help/library/RJafroc/help/SampleSizeGivenJ) | Calculate required number of cases with the given number of readers and DBM variances components. |
| [SaveDataFile](http://127.0.0.1:21213/help/library/RJafroc/help/SaveDataFile) | Save data file in specified format. |
|  |  |

**Author(s)**

Xuetong Zhai, Dev Chakraborty

Maintainer: Xuetong Zhai <xuetong.zhai@gmail.com>

**References**

Chakraborty, D. P., & Berbaum, K. S. (2004). Observer studies involving detection and localization: modeling, analysis, and validation. Medical Physics, 31(8), 1–18. doi:10.1118/1.1769352

[Package *RJafroc* version 1.0 [Index](http://127.0.0.1:21213/help/library/RJafroc/html/00Index.html)]