Package 'SurvHiDim'

January 20, 2025

Title High Dimensional Survival Data Analysis

Type Package

Version 0.1.1

Date 2021-06-22

Description High dimensional time to events data analysis with variable selection technique.
Currently support LASSO, clustering and Bonferroni's correction.

RdMacros Rdpack

LazyDataCompression xz

ByteCompile Yes

License GPL-3

Encoding UTF-8

LazyData true

Depends R (>= 3.5.0)

Imports survival, utils, Rdpack (>= 0.7), readr, tidyverse, glmnet,
factoextra, igraph, useful

Maintainer Atanu Bhattacharjee <atanustat@gmail.com>

RoxygenNote 7.1.1.9000

NeedsCompilation no

Author Atanu Bhattacharjee [aut, cre, ctb],
Akash Pawar [aut, ctb]

Repository CRAN

Date/Publication 2021-06-26 15:40:02 UTC

Contents

hdClust   *hdClust*

Description

Creates a network plot of high dimensional variables and lists those variables.

Usage

```
hdClust(m, n, siglevel, u, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number from where variables of high dimensional data will be selected.
n	Ending column number till where variables of high dimensional data will get selected.
siglevel	Level of significance pre-determined by the user.
u	Factors of Event column e.g. 0,1 or 2 or Number of clusters to form.
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data having survival duration, event information, column of time for death cases and observations on various covariates under study.

Details

Gives network plot and lists the variables showing correlation.

hdClust function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables having P-value < siglevel(significance level taken as input from user) from both subset data. Secondly, it merges the common significant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

hdClust(m,n,siglevel,threshold,data),

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Fress Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - list of variables/genes from deathdata

data2variables - list of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the fucntion stores the output in .csv forms in current directory of user.

Further it creates a cluster plot of variables of similar behavior.

Value

A list containing variable names and the correlation values.

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

See Also

    hdNetwork

Examples

```
##
data(hnscc)
hdClust(7,105,0.05,2,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hdNetwork   *hdNetwork*

Description

    Creates a network plot of high dimensional variables and lists those variables.

Usage

```
hdNetwork(m, n, siglevel, threshold, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
siglevel	Level of significance pre-determined by the user.
threshold	Level to categorize the relation among the significant covariates.
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

    Gives cluster plot and lists the variables showing correlation.

    hdNetwork function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables having P-value < siglevel(significance level taken as input from user) from both subset data. Secondly, it merges the commom sigificant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

hdNetwork(m,n,siglevel,threshold,data),

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Fress Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - list of variables/genes from deathdata

data2variables - list of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the fucntion stores the output in .csv forms in current directory of user.

Further it creates a network plot of variables of similar behavior.

Value

A list containing variable names and the correlation values.

A network plot formed by the filtered variables.

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

See Also

hdClustrr

Examples

```
##
data(hnscc)
hdNetwork(7,105,0.05,0.2,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hidimLasso   *hidimLasso LASSO for high dimensional data*

Description

Least Absolute Shrinkage and Selection Operator (LASSO) for High Dimensional Survival data.

Usage

```
hidimLasso(m, n, OS, Death, data)
```

Arguments

m	Starting column number from where variables of high dimensional data will get selected.
n	Ending column number till where variables of high dimensional data will get selected.
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
data	High dimensional data having survival duration, event and various covariates observations

Details

'HiDimLasso' allows a user to apply LASSO function on the High Dimensional data and reduce the study variables to handful number of co-variate which are observed impacting the survival outcomes.

Column of Overall Survival must be named as 'OS' and the column defining the event must be named as 'Event'.

By default it stores the outcome data in user's current directory.

Value

A list of variables selected by LASSO as predictor variables.

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

See Also

hidimSurvlas hidimSurvbonlas

Examples

```
##
data(hnscc)
hidimLasso(7,105,OS="os",Death="death",hnscc)
##
```

---

hidimSurv   *hidimSurv Survival analysis on high dimensional data*

Description

Survival analysis using Cox Proportional hazards function on high dimensional data

Usage

```
hidimSurv(m, n, siglevel, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
siglevel	Level of significance pre-determined by the user
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"

PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

hidimSurv function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables having P-value < siglevel(significance level taken as input from user) from both subset data. Secondly, it merges the commom sigificant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

hidimSurv(m,n,siglevel,data),

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Fress Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - list of variables/genes from deathdata

data2variables - list of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the fucntion stores the output in .csv forms in current directory of user.

Value

List of variables found significant on OS and survival event

List of variables found significant on PFS and progression event

Estimates values for significant variables on OS and survival event

Estimates values for significant variables on PFS and progression event

Estimates data for the DEGs/Variables found common between significant DEGs from data having death due to progression and data showing death without progression

Estimates data for the DEGs/Variables found common between significant DEGs from data having death due to progression and data showing death without progression

List of variable/ DEGs found common between significantDEGs from data having death due to progression and data showing death without progression

Data with Survival outcomes and DEGs/Variable observations on each subject for DEGs found playing crucial role in death due to progression and without

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

See Also

hidimSurvbon hidimSurvbonlas hidimsvc

Examples

```
##
data(hnscc)
hidimSurv(7,105,0.05,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

| hidimSurvbon | *hidimSurvbon Uses bonferroni correction factor in survival analysis* |

Description

Applies HiDimSurv on high dimensional data using Bonferroni's correction criteria.

Usage

```
hidimSurvbon(m, n, boncorr, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
boncorr	Level of significance on which bonferroni correction will be applied.
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

hidimSurvbon function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables having P-value < siglevel/no. of columns of high dimensional data(significance level taken as input from user) from both subset data. Secondly, it merges the commom sigificant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

hidimSurvbon(m,n,siglevel,data),

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Fress Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - list of variables/genes from deathdata

data2variables - list of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the fucntion stores the output in .csv forms in current directory of user.

Value

List of variables found significant on OS and survival event

List of variables found significant on PFS and progression event

Estimates values for significant variables on OS and survival event

Estimates values for significant variables on PFS and progression event

Estimates data for the DEGs/Variables found common between significant DEGs from data having death due to progression and data showing death without progression

Estimates data for the DEGs/Variables found common between significant DEGs from data having death due to progression and data showing death without progression

List of variable/ DEGs found common between significantDEGs from data having death due to progression and data showing death without progression

Data with Survival outcomes and DEGs/Variable observations on each subject for DEGs found playing crucial role in death due to progression and without

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

See Also

hidimSurvbonlas

Examples

```
##
data(hnscc)
hidimSurvbon(7,105,0.05,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hidimSurvbonlas	*hidimSurvbonlas Two step filteration using bonferroni correction and LASSO*

Description

hidimSurvbonlas on high dimensional data using bonferoni's correction factor and LASSO.

Usage

```
hidimSurvbonlas(m, n, boncorr, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
boncorr	Level of significance on which bonferroni correction will be applied.
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

hidimSurvbonlas function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables using Bonferroni's correction criteria, i.e. having P-value < siglevel/ no. of study variables(significance level taken as input from user) from both subset data. Secondly, it merges the

commom sigificant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

On obtained commondata it fits LASSO and reduces the dimensions of study variables to handful significant variables.

hidimSurvbonlas(m,n,siglevel,data)

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Free Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - list of variables/genes from deathdata

data2variables - list of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the fucntion stores the output in .csv forms in current directory of user.

Value

List of significant genes

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. Political Analysis, 15(3), 199-236. doi: 10.1093/pan/mpl013

See Also

hidimSurvbon

Examples

```
##
data(hnscc)
hidimSurvbonlas(6,104,0.05,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hidimSurvlas   *hidimSurvlas Two step filteration without bonferroni correction*

Description

Survival analysis on high dimensional data using LASSO technique.

Usage

```
hidimSurvlas(m, n, siglevel, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
siglevel	Level of significance pre-determined by the user
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

hidimSurvlas function first creates a new column 'Status' in the input data set and assigns values 0, 1, 2 to each rows. It assigns 0 (for progression = 1 & death(event) = 0) is or (when progression = 0 & death(event) = 0. It assigns 1 (for progression = 1 & death = 1, whereas assigns 2 (for progression = 0 and death = 1).

Further, it creates two data sets, one data set named 'deathdata' which includes subjects with status 0 and 1 and applies Cox PH on it. Another data is named as 'compdata' which includes subjects with status 0 and 2, then applies Cox PH after substituting 2 by 1. Then it filters out study variables having P-value < siglevel(significance level taken as input from user) from both subset data. Secondly, it merges the commom sigificant variables from both data and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant

variables (which are supposed to be leading to death given they leads to progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and correspond in results in .csv format by default in user's current working directory.

On obtained commondata 'hidimSurvlas' fits LASSO and reduces the dimensions of study variables to handful significant variables.

hidimSurvlas(m,n,siglevel,data)

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Free Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

deathdata - A data frame with status column includes only those rows/subjects for which death/event was observed or not, given progression was observed or not.

compdata - A data frame with status column which includes those rows/subjects who died given progression was observed

data1variables - List of variables/genes from deathdata

data2variables - List of variables/genes from compdata

siginificantpvalueA - A data frame with estimate values, HR, Pvalue, etc. of significant variables from deathdata. siginificantpvalueB - A data frame with estimate values, HR, Pvalue, etc. of significant variables from compdata.

commongenes - A data frame consisting observations on common significant study variables.

cvar - List of common significant study variables.

commondata - A final data out consisting survival information ans observations on common significant study variables.

By default the function stores the output in .csv forms in current directory of user.

Value

List of significant genes

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

Examples

```
##
data(hnscc)
hidimSurvlas(7,105,0.05,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hidimsvc	*hidimsvc Survival analysis on high dimensional by creating batches of variables*

Description

Survival analysis on high dimensional data by creating batches of covariates

Usage

```
hidimsvc(m, n, batchsize, siglevel, ID, OS, Death, PFS, Prog, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
batchsize	Number of variables to be consider at time while running function (maximum batch size should not be greater than one third of the total number of high dimensional variables)
siglevel	Level of significance pre-determined by the user
ID	Column name of subject ID, a string value. i.e. "id"
OS	Column name of survival duration event, a string value. i.e. "os"
Death	Column name of survival event, a string value. i.e "death"
PFS	Column name of progression free survival duration, a string value. i.e "pfs"
Prog	Column name of progression event, a string value. i.e "prog"
data	High dimensional data containing the survival, progression and genomic observations.

Details

hidimsvc function fits Univarate Cox Proportinal Hazard models by considering each variables at a time. Then it filters out study variables having P-value < siglevel(significance level taken as input from user). Once by survival and survival eevent and another by progression and progression events. Secondly, it merges the commom sigificant variables from both OS and PFS analysis and creates a new data frame which contains columns, 'ID','OS','Death','PFS','Prog','Status' and observations of common significant variables (which are supposed to be leading to death given they leads to

progression of cancer as well as accounts for competing risks) Further, it lists the common variables names and outs corresponding results in .csv format by default in user's current working directory.

It works similary to HiDimSurv unlike it creates batches of decided study vriables by user to make the analysis less time consuming.

hidimsvc(m1,m2,batchsize,siglevel,data),

1) Subject ID column should be named as 'ID'.

2) OS column must be named as 'OS'.

3) Death status/event column should be named as 'Death'.

4) Progression Fress Survival column should be named as 'PFS'.

5) Progression event column should be named as 'Prog'.

OSDeathcoeff - A data frame containing HR estimates and p-values for study variables on fitting univariate CoxPh on OS and Survival event.

PFSProgcoeff - A data frame containg HR estimates and p-values for study variables on fitting univariate CoxPh on PFS and Progression event.

namevect - List of all the study variable names.

significantOSDeathgenes - A data frame containing HR estimates and p-values for significant study variables.

significantPFSProggenes - A data frame containing HR estimates and p-values for significant study variables

commongenes - A data frame containing estimated values of significant study variables found common from significant study variables on fitting CoxPh on survival and progression times and events.

odnames - List of significant variables on fitting CoxPh using survival and survival event.

ppnames - List of significant variables on fitting CoxPh using progression and progression event.

cvar - List of common significant study variables on fitting CoxPh on survival and progression, times and events.

commondata - A data out which contains the clinical observations and observations on common-genes variables.

Value

Estimate values of significant variables/DEGs on considering Death with Progression

Estimate values of significant variables/DEGs on considering Death without Progression

List of variable/DEGs considering Death with Progression

List of variable/DEGs considering Death without Progression

Estimates data for the DEGs/Variables found common between significant DEGs from data having death due to progression and data showing death without progression"

List of variable/DEGs found common between significant DEGs from data having death due to progression and data showing death without progression"

Author(s)

Atanu Bhattacharjee and Akash Pawar

References

Bhattacharjee, A. (2020). Bayesian Approaches in Oncology Using R and OpenBUGS. CRC Press.

Congdon, P. (2014). Applied bayesian modelling (Vol. 595). John Wiley & Sons.

Banerjee, S., Vishwakarma, G. K., & Bhattacharjee, A. (2019). Classification Algorithm for High Dimensional Protein Markers in Time-course Data. arXiv preprint arXiv:1907.12853.

Examples

```
##
data(hnscc)
hidimsvc(7,105,5,0.05,ID="id",OS="os",Death="death",PFS="pfs",Prog="prog",hnscc)
##
```

---

hnscc   *hnscc High dimensional genomic data on head and neck cancer*

Description

High dimensional breast cancer gene expression data

Usage

```
hnscc
```

Format

A dataframe with 565 rows and 104 variables

id ID of subjects

leftcensor Initial censoring time

death Survival event

os Duration of overall survival

pfs Duration of progression free survival

prog Progression event

GJB1,...,HMGCS2 High dimensional covariates

Examples

```
## Not run:
data(hnscc)

## End(Not run)
```

---

multicoxa   *High dimensional multivariate cox proportional hazard data analysis*

Description

Given the dimensions of the variables and survival informations. The function performs multivariate Cox PH by taking 5 variables at a time.

Usage

```
multicoxa(m, n, OS, event, data)
```

Arguments

m	Starting number of column from where multivariate variables will get selected.
n	Ending number of column till where multivariate variables will get selected.
OS	"Column/Variable name" consisting duration of survival.
event	"Column/Variable name" consisting survival event.
data	High dimensional data containing survival observations and covariates.

Value

Data set containing the survival estimates and Pvalue.

Examples

```
##
multicoxa(m=15,n=18,OS="os",event="death",data=hnscc)
```

---

multicoxb   *High dimensional multivariate cox proportional hazard data analysis*

Description

Given the dimensions of the variables and survival informations. The function performs multivariate Cox PH by taking 5 variables at a time.

Usage

```
multicoxb(
  C1 = NULL,
  C2 = NULL,
  C3 = NULL,
  C4 = NULL,
  C5 = NULL,
  OS,
  event,
  data
)
```

Arguments

C1	Covar1
C2	Covar2
C3	Covar3
C4	Covar4
C5	Covar5
OS	"Column/Variable name" consisting duration of survival.
event	"Column/Variable name" consisting survival event.
data	High dimensional data containing survival observations and covariates.

Value

Data set containing the survival estimates and Pvalue.

Examples

```
##
multicoxb(C1="GJB1",C2=NULL,C3="HPN",C4=NULL,C5=NULL,OS="os",event="death",data=hnscc)
```

---

survdesc	*High dimensional univariate cox proportional hazard analysis.*

Description

Given the dimension of variables and survival information the function performs uni-variate Cox PH.

Usage

```
survdesc(m, n, survdur, event, aic = TRUE, data)
```

Arguments

m	Starting column number form where study variables of high dimensional data will get selected.
n	Ending column number till where study variables of high dimensional data will get selected.
survdur	Column name of survival duration event, a string value. i.e. "os"
event	Column name of survival event, a string value. i.e "death"
aic	By default aic = FALSE, if aic = TRUE the function returns
data	High dimensional data containing the survival, progression and genomic observations.

Value

A data set containing estimates for variables present in column m to n.

Examples

```
##
data(hnscc)
survdesc(m=10,n=50,survdur="os",event="death",aic=TRUE,data=hnscc)
##
```

Index