Supporting Information for "Influence Analysis in Quantitative Trait Loci Detection" by Dou et al. — Software Description

EIFs.R

Description

The R module `EIFs.R` contains two functions `EIFs()` and `plot.eifs()`, and an example that uses the dataset from the paper.

`EIFs()` returns (1) a vector of LOD scores for all marker loci, (2) a matrix containing the empirical influence functions for all marker loci, (3) a vector of linear combinations of empirical influence functions, and (4) a matrix of empirical influence functions for specified marker loci. This function computes `Eifc` for an arbitrarily chosen coefficient vector and marker loci.

`plot.eifs()` plots the vector of the `Eifc` obtained from `EIFs()`.

Usage

```
EIFs(X, loci, cl=(1,1,1,1))
plot.eifs(eifc, mx=1)
```

Arguments

`X` a matrix of the dataset.

`loci` a vector to specify marker loci.

`cl` a coefficient vector of empirical influence functions corresponding to the specified marker loci.

`eifc` a vector obtained from `EIFs()`.

`mx` how many of the largest values of `eifc` to show in the figure.

Details

`X` contains ID, phenotype, sex and genotypes. The first row is the description of each column, the second row contains the names of chromosomes, and the third row contains the locations of marker loci. In the rest of the dataset, each row consists of an ID number, a phenotype value, a sex indicator (0 for female or 1 for male), and genotypes for the marker loci.

`loci` consists of integers no larger than the total number of marker loci.

`cl` should be the same length as `loci`.

`mx` is the number of most influential individuals whose ID number and EIF values will be shown in the figure. Its default value is 1.

Example

```
# dataset #
X1 <- matrix(scan("Data_LogADI.txt", what="character"), 173, byrow=TRUE)


# Input the dataset, specify marker loci and provide a coefficient vector
#  cl to the function EIFs.  #
eifc<- EIFs(X=X1, loci=c(2,30,110), cl=c(0.2,2,0.5))


# Function EIFs() returns LOD scores, EIF for all loci, and Eifc for the
#  specified loci.  #
Eifc<- eifc$Eifc


# Plot Eifc and show the mx individuals having the largest absolute Eifc
#  values.  #
plot.eifs(Eifc, mx=3)
```

Orthogonal.R

Description

The R module `Orthogonal.R` contains two functions `orthogonal.f()` and `plot.eifc()`, and an example that uses the dataset from the paper.

`orthogonal.f()` returns (1) a vector of LOD scores for all marker loci, (2) a matrix containing the empirical influence functions for all marker loci, (3) a matrix of linear combination of empirical influence functions, (4) a list of orthogonal-projection matrices, and (5) a matrix of the orthogonal coefficient vectors. This function computes `Eifc` for specified marker loci on the same chromosome. Orthogonal projection matrices and coefficient vectors are computed by Gram-Schmidt orthonormalization.

`plot.eifc()` plots the vector of `Eifc` obtained from `orthogonal.f()`. The degree of the orthogonal polynomial should be specified.

Usage

```
orthogonal.f(X, chromosome, loci, power)
plot.eifs(eifc, p, mx=1)
```

Arguments

`X` a matrix of the dataset.

`chromosome` an integer number or a character "X" to specify a chromosome.

`loci` a vector to specify marker loci.

`power` the degree of the orthogonal polynomial.

`eifc` a vector obtained from `orthogonal.f()`.

`p` an integer no larger than `power`.

`mx` how many of the largest values of `eifc` to show in the figure.

Details

`X` contains ID, phenotype, sex, and genotypes. The first row is the description of each column, the second row contains the names of chromosomes, and the third row contains the locations of marker loci. In the rest of the dataset, each row consists of an ID number, a phenotype value, a sex indicator (0 for female or 1 for male), and genotypes for the marker loci.

`loci` consists of integers no larger than the total number of marker loci on the specified chromosome.

`mx` is the number of most influential individuals whose ID number and EIF values will be shown in the figure. Its default value is 1.

Example

```
# dataset #
X1 <- matrix(scan("Data_LogADI.txt", what="character"), 173, byrow=TRUE)


# Input the dataset and specify the chromosome, the marker loci and the
#  power of the orthogonal polynomial.  #
ortho.eifc<- orthogonal.f(X=X1, chromosome=3, loci=c(3,4,5), power=2)
EIFC <- ortho.eifc$Eifc


# Plot EIFC, specify p (any integer from 0 to the power), and show the mx
#  individuals having the largest absolute EIFCs.  #
plot.eifc(EIFC, p=2, mx=3)
```

PCA_scores.R

Description

The R module `PCA_score.R` contains three functions `ISVs()`, `plot.isv()` and `plot2d.isv()`, and an example that uses the dataset from the paper.

`ISVs()` returns (1) a vector of LOD scores for all marker loci, (2) a matrix containing the empirical influence functions for all marker loci, (3) a vector of eigenvalues, (4) a matrix of eigenvectors, (5) a matrix of influence score vectors, and (6) a vector of positive eigenvalues.

`plot.eifc()` plots the influence score vector obtained from `ISVs()`. The index for the influence score vector should be specified.

`plot2d.isv()` plots two influence score vectors obtained from `ISVs()`. The index of each influence score vector should be specified.

Usage

```
ISVs(X, chromosome, loci)
plot.isv(ISV, d, mx=1)
plot2d.isv(ISV, d1, d2, mx=1)
```

Arguments

`X` a matrix of the dataset.

`chromosome` an integer number or a character "X" to specify a chromosome.

`loci` a vector to specify marker loci.

`ISV` a matrix obtained from `ISVs()`.

`d, d1, d2` positive integers no larger than `ncol(ISV)`.

`mx` how many of the largest EIF values to show in the figure.

Details

`X` contains ID, phenotype, sex, and genotypes. The first row is the description of each column, the second row contains the names of chromosomes, and the third row contains the locations of marker loci. In the rest of the dataset, each row consists of an ID number, a phenotype value, a sex indicator (0 for female or 1 for male), and genotypes for the marker loci.

`loci` consists of integers not larger than the total number of marker loci on the specified chromosome.

`mx` is the number of most influential individuals whose ID number and EIF values will be shown in the figure. The default value is 1.

Example

```
# dataset #
X1 <- matrix(scan("Data_LogADI.txt", what="character"), 173, byrow=TRUE)

# Input the dataset and specify chromosome and marker loci #
isvs<- ISVs(X=X1, chromosome=3, loci=c(3,4,5))
isv <- isvs$ISV

# Plot isv #
# Indicate the index(es) d (or d1, d2) for the influence score vector(s) #
plot.isv(ISV=isv, d=2, mx=3)
plot2d.isv(ISV=isv, d1=1, d2=2, mx=3)
```

QEIF.R

Description

The R module `QEIF.R` contains two functions `QEIFs()` and `plot.qeif()` and an example that uses the dataset from the paper.

`QEIFs()` returns (1) a vector of LOD scores for all marker loci, (2) a matrix containing the empirical influence functions for all marker loci, (3) a matrix of orthogonal projection matrix, and (4) a vector of the quadratic empirical influence functions for all individuals.

`plot.qeif()` plots the vector of `qeif` obtained from `QEIFs()`.

Usage

```
QEIFs(X, chromosome, loci)
plot.qeif(qeif, mx=3)
```

Arguments

`X` a matrix of the dataset.

`chromosome` an integer number or a character "X" to specify a chromosome.

`loci` a vector to specify marker loci.

`qeif` a vector obtained from `QEIFs()`.

`mx` the number of largest values of `qeif` to show in the figure.

Details

`X` contains ID, phenotype, sex, and genotypes. The first row is the description of each column, the second row contains the names of chromosomes, and the third row contains the locations of marker loci. In the rest of the dataset, each row consists of an ID number, a phenotype value, a sex indicator (0 for female or 1 for male), and genotypes for the marker loci.

`loci` consists of integers not larger than the total number of marker loci on the specified chromosome.

`mx` is the number of most influential individuals whose ID number and EIF values will be shown in the figure. Its default value is 3.

Example

```
# dataset #
X1 <- matrix(scan("Data_LogADI.txt", what="character"), 173, byrow=TRUE)


# Input the dataset and specify chromosome and marker loci.  #
qeifs<- QEIFs(X=X1, chromosome=3, loci=c(3,4,5))
QEIF <- qeifs$Qeif


# Plot quadratic EIFs #
plot.qeif(QEIF, mx=3)
```