Package 'evian'

May 6, 2010

Type Package

Title Evidential Analysis of Genetic Association Data

Version 1.1.0

Date 2010-05-06

Author Dr. Lisa J Strug <lisa.strug@utoronto.ca>

Maintainer Ted Chiang <tchiang@sickkids.ca>

Description Evidential regression analysis for dichotomous and quantitative outcome data.

License GPL (>=2)

URL http://strug.ccb.sickkids.ca/evian

R topics documented:

evian-package *Evidential Analysis of Genetic Association Data*

Description

The evidential analysis approach employs the law of likelihood, and uses the likelihood ratio instead of p-values to analyze and interpret genetic association data. The current package analyzes dichotomous and quantitative outcome data.

Details

Package:	evian
Type:	Package
Version:	1.0
Date:	2010-05-06
License:	GPL (>= 2)

The main evian functions are `evian_logit` and `evian_linear` which uses logistic and linear regression models respectively. It can accomodate covariates and code the SNP marker as dominant, recessive, additive, 2df overdominance, or 2df genotypic.

Author(s)

Dr. Lisa J Strug <lisa.strug@utoronto.ca>

Maintainer: Ted Chiang <tchiang@sickkids.ca>

References

Strug, Rohde & Corey. Am Stat, 61:207-212; Strug & Hodge, I. Human Heredity, 61:166-188; Strug & Hodge, II. Human Heredity, 61:200-209; Strug et. al, 2009. EJHG, 17:1171-1181; Strug et. al, 2010. EJHG, doi:10.1038,ejhg.2010.47;

See Also

`evian_logit`, `evian_logit_plotsnp`, `logit_snp`, `logit_plotsnp li_plot`, `evian_linear`, `evian_linear_plotsnp`, `linear_snp`, `linear_plotsnp`

---

eviandata *An example genotype and covariate dataset to illustrate evian functions for logistic regression*

Description

The 'evian_data' data frame has 250 rows and 39 columns of improvised snp genotype and covariate data imitating individuals in a genetic association study.

Usage

```
eviandata
```

Format

This data frame contains the following columns. The first six columns are analogous to the LINKAGE format, followed by three columns contanining covariate data: age, weight, city. And finally thirty columns of snp genotype data: 0, 1, 2, and NA, which code for the number of minor alleles in the genotype call.

FID is an identifier for the family.
IID is an identifier for the individual.
PAT is an identifier for the IID's father.
MAT is an identifier for the IID's mother.
SEX is a numeric for the IID's gender 1 = M, 2 = F.
PHENOTYPE is a numeric for the case status for the IID, 0 = unaffected, 1 = affected.
age is a numeric for the IID's age.
weight is a numeric for the IID's weight.
city is a factor with levels 'toronto', 'montreal', 'vancouver' for the IID's city.
snp genotype data.

---

eviandata_linear *An example genotype and covariate dataset to illustrate evian functions for linear regression*

Description

The 'eviandata_linear' data frame has 1444 rows and 19 columns of improvised snp genotype and covariate data imitating individuals in a genetic association study.

Usage

```
eviandata_linear
```

Format

This data frame contains the following columns. The first six columns are analogous to the LINKAGE format, followed by three columns contanining covariate data: Fev, BMI_group, Age_group. And finally ten columns of snp genotype data: 0, 1, 2, and NA, which code for the number of minor alleles in the genotype call.

FID is an identifier for the family.
IID is an identifier for the individual.
PAT is an identifier for the IID's father.
MAT is an identifier for the IID's mother.
SEX is a numeric for the IID's gender 1 = M, 2 = F.
Y_norma is a continuous variable for the Y outcome data.
Fev is a continuous covariate.
BMI_group is a categorical covariate for the IID's BMI.
Age_group is a categorical covariate for the IID's age.
snp genotype data.

evianmap — *An example snp map to illustrate evian functions for logistic regression*

Description

The 'evianmap' data frame has 30 rows and 3 columns of improvised snp map data used in genetic association study.

Usage

```
evianmap
```

Format

This data frame contains the following three columns.

chr is a numeric indicating the chromosome number.
snp is a string indicating the snp name.
pos is a numeric indicating the snp base position on the chromosome.

evianmap_linear — *An example snp map to illustrate evian functions for linear regression*

Description

The 'evianmap_linear' data frame has 10 rows and 3 columns of improvised snp map data used in genetic association study.

Usage

```
evianmap_linear
```

Format

This data frame contains the following three columns.

chr is a numeric indicating the chromosome number.
snp is a string indicating the snp name.
pos is a numeric indicating the snp base position on the chromosome.

evian_linear *Evidential Analysis for quantitative outcome data using linear regression models*

Description

Calculate and plot the likelihood intervals for genetic association of a quantitative trait in a genomic region of interest. Covariates can be accommodated.

Usage

```
evian_linear(data, map, ycol, xcols, formula, robust = FALSE,
             model = 'all', m = 1000, bse = 5,
             lolim, hilim, plot_li = TRUE, yaxislim = 5,
             showmaxlr = 3, kcutoff = 32, grays = TRUE,
             save_as_png = FALSE)
```

Arguments

`data`	a data frame whose first six columns are in standard linkage format: FID, IID, PAT, MAT, SEX `1 = M, 2 = F`, Y_OUTCOME, followed by columns of non-genetic covariates, and then the columns of genotype data coded `0,1,2,NA` as produced in the ped file by the plink software with option `--recodeA`.
`map`	a data frame with three column header: chr, snp, pos, ie. map file in plink format.
`ycol`	numeric, the column number in the `data` data frame for the Y outcome variable.
`xcols`	range, the column range in the `data` data frame for the snp data, ie. `8:17`.
`formula`	string containing the glm formula for Y, e.g., `Y ~ x + Z$age` for models where the snp covariate is for BOTH 1df and 2df snp parameterization. The additional term `x1` is not required nor permitted for 2df models. Use of `as.factor(Z$var)` for categorical variable is recommended. Note, the formula must be at least `Y ~ x` for 1df and 2df.
`robust`	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
`model`	a string specifying the mode of inheritance parameterization: `additive, dominant, recessive, overdominance, 2df` or `all`.
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e. default = `5`.
`lolim`	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the regression parameter B, at which to calculate the likelihood function.
`hilim`	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the regression parameter B, at which to calculate the likelihood funciton.
`plot_li`	logical, produce the likelihood intervals plot in a graphics window (X11 or Quartz).

yaxislim	y axis limit for the beta, default = 5.
showmaxlr	specifies the number of SNPs for which the user wants the maximum likelihood ratios displayed in text on the plot, default = 3.
kcutoff	the strength of evidence criterion k: 32, 100, 1000, default = 32.
grays	logical, if TRUE, then individual intervals for a given SNP will be in gray if the 1/k likelihood interval includes beta = 0 as a plausible value.
save_as_png	logical, if TRUE the likelihood intervals plot will be saved as a png file. Filename will be a time stamp in current directory.

Details

linear_snp is main function called to calculate the 1/k likelihood intervals for the additive, dominant, recessive, overdominance and 2df genotypic models.

Usually the data inputs, ie. data and map, are conformed to the plink format when run with the --recodeA option.

The formula must be at least Y ~ x for snp 1df or 2df models. The additional term x1 is not required or permitted for 2df models. Exact variable syntax and character spacing in formula is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in formula.

Value

This function produces a data frame named according to the selected model, ie. data_linear.additive, with the following 10 column headings:

snp, pos, maxlr, mle, k32_low, k32_hi, k100_low, k100_hi, k1000_low, k1000_hi

By default, the likelihood intervals are plotted in a graphics window (X11 or Quartz), and/or the image is saved in png file if save_as_png is TRUE. In the plot, the SNPs with beta = 0 in the 1/k likelihood interval will be in gray if grays = TRUE. Those with beta >= 0 not in the 1/k likelihood interval will be in color with the 1/32 intervals in green, 1/100 in red and 1/1000 in blue. maxLR is defined as

$$L(\hat{\theta})/L(0)$$

, where $\hat{\theta}$ is the maximum likelihood estimate of theta.

Rename the data frame in your R workspace before running another analysis to prevent overwriting results.

Note

When lolim/hilim is NOT defined, then the boundaries of the beta grid will be determined by the default bse=5, or a bse defined by the user. Otherwise, the user can define the exact beta grid boundaries using lolim/hilim.

In some cases the beta grid (using bse or lolim/hilim) may need to be increased substantially (bse as large as 15) if covariates are present in the formula.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Blume JD, et. al. Stat Med. 26(15):2919-36.

See Also

evian_linear_plotsnp, li_plot, linear_snp, linear_plotsnp

Examples

```
## Not run:
evian_linear(data, map, ycol, xcols, formula, robust = FALSE,
            model = 'all', m = 1000, bse = 5,
            lolim, hilim, plot_li = TRUE, yaxislim = 5,
            showmaxlr = 3, kcutoff = 32, grays = TRUE,
            save_as_png = FALSE)

## End(Not run)
```

---

evian_linear_plotsnp

Plotting likelihoods for single SNPs for quantitative outcome data

---

Description

For a single snp, this function produces a standardized likelihood curve for genetic association of a quantitative trait: beta on the x-axis, against standardized likelihood on the y-axis.

Usage

```
evian_linear_plotsnp (data, map, ycol, xcols, formula,
                      robust = FALSE, snp, model = '',
                      m = 1000, bse = 5, lolim, hilim,
                      k1 = 32, k2 = 100,
                      print.details = TRUE)
```

Arguments

data	a data frame whose first six columns are in standard linkage format: FID, IID, PAT, MAT, SEX 1 = M, 2 = F, Y_OUTCOME, followed by columns of non-genetic covariates, and then the columns of genotype data coded 0,1,2,NA as produced in the ped file by the plink software with option --recodeA.
map	a data frame with three column header: chr, snp, pos, ie. map file in plink format.
ycol	numeric, the column number in the data data frame for the Y outcome variable.
xcols	range, the column range in the data data frame for the snp data, ie. 8:17.
formula	string containing the glm formula for Y, e.g., Y ~ x + Z$age for models where the snp covariate is for BOTH 1df and 2df snp parameterization. The additional term x1 is not required nor permitted for 2df models. Use of as.factor(Z$var) for categorical variable is recommended. Note, the formula must be at least Y ~ x for 1df and 2df.

robust	logical, if TRUE, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
snp	string, the snp name of interest stored in the map data frame
model	a string specifying the mode of inheritance parameterization: additive, dominant, recessive, overdominance, 2df or all.
m	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
bse	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e. default = 5.
lolim	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the regression parameter B, at which to calculate the likelihood function.
hilim	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the regression parameter B, at which to calculate the likelihood funciton.
k1	numeric, 1/k1 likelihood interval: 8, 32, 100, 1000, plotted in green
k2	numeric, 1/k2 likelihood interval: 8, 32, 100, 1000, plotted in red
print.details	logical, if TRUE, then displays plot legend and details

Details

This function will produce the likelihood function plot for a linear regression model.

The formula must be at least Y ~ x for snp 1df or 2df models. The additional term x1 is not required or permitted for 2df models. Exact variable syntax and character spacing in formula is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in formula.

Usually the data inputs, ie. data and map, are conformed to the plink format when run with the --recodeA option.

Value

This function will produce a plot in a X11 window or a Quartz device window for MacOSX.

Note

When lolim/hilim is NOT defined, then the boundaries of the beta grid will be determined by the default bse=5, or a bse defined by the user. Otherwise, the user can define the exact beta grid boundaries using lolim/hilim.

The beta grid limits may need to be adjusted or broadened in plots when the calculated likelihood intervals are not available, e.g., NAs provided.

In some cases the beta grid (using bse or lolim/hilim) may need to be increased substantially (bse as large as 15) if covariates are present in the formula.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Enable the X11 server on your terminal before running R. If your desktop OS is Linux or MacOSX (10.5), the X11 server should be automatic. In MS-Windows, you may need launch an X server appliction such as Xming.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Blume JD, et. al. Stat Med. 26(15):2919-36.

See Also

evian_linear_plotsnp, li_plot, linear_snp, linear_plotsnp

Examples

```
## Not run:
evian_linear_plotsnp (data, map, ycol, xcols, formula,
                      robust = FALSE, snp, model = '',
                      m = 1000, bse = 5, lolim, hilim,
                      print.details = TRUE)

## End(Not run)
```

---

evian_logit — *Evidential Analysis for dichotomous outcome data using logistic regression models*

Description

Calculate and plot the likelihood intervals for genetic association in a genomic region of interest with a dichotomous trait. Covariates can be accommodated.

Usage

```
evian_logit(data, map, ycol, xcols, formula, robust = FALSE,
            model = 'all', rfdum = 'alphanum', m = 200, bse,
            lolim = log(0.025), hilim = log(4),
            plot_li = TRUE, yaxislim = 10, showmaxlr = 3,
            kcutoff = 32, grays = TRUE, save_as_png = FALSE)
```

Arguments

`data`	a data frame whose first six columns are in standard linkage format: FID, IID, PAT, MAT, SEX `1 = M, 2 = F`, PHENOTYPE `0,1`, followed by columns of non-genetic covariates, and then the columns of genotype data coded `0,1,2,NA` as produced in the ped file by the plink software with option `--recodeA`.
`map`	a data frame with three column header: chr, snp, pos, ie. map file in plink format.
`ycol`	numeric, the column number in the `data` data frame for the Y outcome variable, ie. PHENOTYPE or caseflag.
`xcols`	range, the column range in the `data` data frame for the snp data, ie. `8:17`.

formula	string containing the glm formula for Y, e.g., `Y ~ x + Z$age` for models where the snp covariate is a 1df parameterization, or `Y ~ x + x1 + Z$age` for 2df snp parameterization. Use `as.factor(Z$var)` if the variable is categorical. Note, the formula must be at least `Y ~ x` for 1df, and `Y ~ x + x1` for 2df.
robust	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
model	a string specifying the mode of inheritance parameterization: `additive, dominant, recessive, overdominance, 2df` or `all`.
rfdum	string to designate the reference group for categorical variables. Options: `alphanum` orders levels alphanumerically, `most` chooses the level observed most frequently, `least` chooses the level observed least frequently.
m	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
bse	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e.
lolim	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the log(OR), at which to calculate the likelihood function.
hilim	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the log(OR), at which to calculate the likelihood funciton.
plot_li	logical, produce the likelihood intervals plot in graphics window (X11 or Quartz).
yaxislim	y axis limit for the odds ratio, default = `10`.
showmaxlr	specifies the number of SNPs for which the user wants the maximum likelihood ratios displayed in text on the plot, default = `3`.
kcutoff	the strength of evidence criterion k: `32, 100, 1000`, default = `32`.
grays	logical, if `TRUE`, then individual intervals for a given SNP will be in gray if the `1/k` likelihood interval includes `OR = 1` as a plausible value.
save_as_png	logical, if `TRUE` the likelihood intervals plot will be saved as a png file. Filename will be a time stamp in current directory.

Details

`logit_snp` is main function called to calculate the `1/k` likelihood intervals for the additive, dominant, recessive, overdominance and 2df genotypic models.

Usually the data inputs, ie. data and map, are conformed to the plink format when run with the `--recodeA` option.

The `formula` must be at least `Y ~ x` for snp 1df models, and `Y ~ x + x1` for snp 2df models. Exact variable syntax and character spacing in `formula` is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in `formula`.

Value

This function produces a data frame named according to the selected model, ie. data.additive, with the following 11 column headings:

snp, pos, maxlr, mle, k32_low, k32_hi, k100_low, k100_hi, k1000_low, k1000_hi, flip

When the value of mle is less than one, the k intervals are inverted and flip=1, otherwise flip=0.

By default, the likelihood intervals are plotted in an X11 window, and/or the image is saved in png file if `save_as_png` is `TRUE`. In the plot, the SNPs with `OR = 1` in the `1/k` likelihood interval will be in gray if `grays = TRUE`. Those with `OR >= 1` not in the `1/k` likelihood interval will be in color with the 1/32 intervals in green, 1/100 in red and 1/1000 in blue. `maxLR` is defined as

$$L(\hat{\theta})/L(1)$$

, where $\hat{\theta}$ is the maximum likelihood estimate of the log(OR).

Rename the data frame in your R workspace before running another analysis to prevent overwriting results.

Note

When the number of beta standard errors `bse` is NOT defined, then the boundaries of the beta grid will use the default `lolim/hilim` values, or those defined by the user. However, if `bse` is defined, then `lolim/hilim` values will be ignored.

In some cases the beta grid (using `bse` or `lolim/hilim`), may need to be increased substantially (`bse` as large as 15) if covariates are present in the formula.

Depending on the size of the dataset, this function may require long compute times. A typical runtime for 1df models, ie. additive, dominant or recessive, without any covariates on a single processor of a 2.53 GHz Intel Core 2 Duo will be around 3 minutes for a dataset containing 50 snps with 250 individuals.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Strug et. al, 2009. EJHG, 17:1171-1181;

See Also

`evian_logit_plotsnp`, `li_plot`, `logit_snp`, `logit_plotsnp`

Examples

```
## Not run:
evian_logit(data, map, ycol, xcols, formula, model = "all",
            rfdum = 'alphanum', m = 200, bse, lolim = log(0.025),
            hilim = log(4), plot_li = TRUE, yaxislim = 10,
            showmaxlr = 3, kcutoff = 32, grays = TRUE,
            save_as_png = FALSE)

## End(Not run)
```

`evian_logit_plotsnp`

Plotting likelihoods for single snps

Description

For a single snp, this function produces a standardized likelihood curve: OR on the x-axis, against standardized likelihood on the y-axis.

Usage

```
evian_logit_plotsnp (snp, data, map, ycol, xcols, formula,
                     robust = FALSE, model ='',
                     rfdum='alphanum', m=200,
                     bse, lolim=log(0.025), hilim=log(4),
                     k1=32, k2=100, print.details=TRUE)
```

Arguments

`snp`	string, the snp name of interest stored in the `map` data frame, e.g. rs21547.
`data`	a data frame whose first six columns are in standard linkage format: FID, IID, PAT, MAT, SEX `1 = M, 2 = F`, PHENOTYPE `0,1`, followed by columns of non-genetic covariates, and then the columns of genotype data coded `0,1,2,NA` as produced in the ped file by the plink software with option `--recodeA`.
`map`	a data frame with three column header: chr, snp, pos, ie. map file in plink format.
`ycol`	numeric, the column number in the `data` data frame for the Y outcome variable, ie. PHENOTYPE or caseflag.
`xcols`	range, the column range in the `data` data frame for the snp data, ie. `8:17`.
`formula`	string containing the glm formula for Y, e.g., `Y ~ x + Z$age` for models where the snp covariate is a 1df parameterization, or `Y ~ x + x1 + Z$age` for 2df snp parameterization. Use `as.factor(Z$var)` if the variable is categorical. Note, the formula must be at least `Y ~ x` for 1df, and `Y ~ x + x1` for 2df.
`robust`	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
`model`	a string specifying the mode of inheritance parameterization: `additive, dominant, recessive, overdominance, 2df` or `all`.
`rfdum`	string to designate the reference group for categorical variables. Options: `alphanum` orders levels alphanumerically, `most` chooses the level observed most frequently, `least` chooses the level observed least frequently.
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e.
`lolim`	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the log(OR), at which to calculate the likelihood function.

hilim	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the log(OR), at which to calculate the likelihood funciton.
k1	numeric, 1/k1 likelihood interval: 8, 32, 100, 1000, plotted in green
k2	numeric, 1/k2 likelihood interval: 8, 32, 100, 1000, plotted in red
print.details	logical, if TRUE, then the plot is displayed with legend and details

Details

This function will produce the likelihood function plot for a logistic regression model.

The formula must be at least Y ~ x for snp 1df models, and Y ~ x + x1 for snp 2df models. Exact variable syntax and character spacing in formula is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in formula.

Value

This function will produce a plot in a X11 window or a Quartz device window for MacOSX.

Note

When the number of beta standard errors, bse is NOT defined, then the boundaries of the beta grid will use the default lolim/hilim values, or those defined by the user. However, if bse is defined, then lolim/hilim values will be ignored.

The beta grid limits may need to be adjusted or broadened in plots when the calculated likelihood intervals are not available, e.g., NAs provided.

In some cases the beta grid (using bse or lolim/hilim), may need to be increased substantially (bse as large as 15) if covariates are present in the formula.

For a given plot, the typical runtime for 1df models, ie. additive, dominant or recessive, without any covariates on a single processor of a 2.53 GHz Intel Core 2 Duo will be around 5 seconds for a dataset with 250 individuals.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Enable the X11 server on your terminal before running R. If your desktop OS is Linux or MacOSX (10.5), the X11 server should be automatic. In MS-Windows, you may need launch an X server appliction such as Xming.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Royall, R.M,. Statistical evidence: a likelihood paradigm. ISBN:9780412044113. Edwards, A.W.F., Likelihood ISBN:0801844436. Strug & Hodge, I. Human Heredity, 61:166-188; Strug & Hodge, II. Human Heredity, 61:200-209; Blume JD, et. al. Stat Med. 26(15):2919-36.

See Also

evian_logit, li_plot, logit_snp, logit_plotsnp

Examples

```
## Not run:
evian_logit_plotsnp (snp, data, map, ycol, xcols, formula,
                     model ='', rfdum='alphanum', m=200,
                     bse, lolim=log(0.025), hilim=log(4),
                     k1=32, k2=100, print.details=TRUE)

## End(Not run)
```

`linear_plotsnp` *Linear regression likelihood plotting function called in evian_linear_plotsnp*

Description

This is the core linear likelihood function that plots the likelihood function for a single SNP. This function is used for the additive, dominant, recessive, 2df overdominance and genotypic models.

Usage

```
linear_plotsnp(X, Y, Z, cluster.id, userformulaX,
               userformulaY, snp,  m=1000, bse=5,
               lolim, hilim, robust=FALSE,
               k1=32, k2=100, print.details=TRUE)
```

Arguments

`X`	a vector of snp genotype data coded, `0,1,2,NA`
`Y`	a vector of the continuous outcome variable
`Z`	a data frame with the first column containing the "X intercept" $vectorof1's$, and followed by columns of x predictor data $covariates$. The header for the first column of intercept must be 'intercept' because `Z$intercept` is used in the variables `userformulaX, userformulaY` .
`cluster.id`	a vector of family ids, that groups a number of related individuals.
`userformulaX`	string containing the glm formula for X, ie. `X ~ Z$intercept + Z$weight` . Use `as.factor(Z$var)` if the data is categorical.
`userformulaY`	string containing the glm formula for Y, ie. `Y ~ -1 + Z$intercept + Z$weight`. Use `as.factor(Z$var)` if the data is categorical.
`snp`	string, the snp name of interest stored in the map data frame
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e. default = `5`.
`lolim`	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the regression parameter B, at which to calculate the likelihood function.

hilim	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the regression parameter B, at which to calculate the likelihood funciton.
robust	logical, if TRUE, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. cluster.id, or FID.
k1	numeric, 1/k1 likelihood interval: 8,32,100,1000, plotted in green
k2	numeric, 1/k2 likelihood interval: 8,32,100,1000, plotted in red
print.details	logical, if TRUE, then displays plot legend and details

Details

This core function is run iteratively in a loop within evian_linear for the following selected models: additive, dominant, recessive, 2df overdominance and genotypic.

Depending on which 1df model is desired, the SNP data X will need to be coded accordingly. Coding should be as follows:

```
   Additive
 AA  0
 AB  1
 BB  2

   Dominant
 AA  0
 AB  1
 BB  1

   Recessive
 AA  0
 AB  0
 BB  1
```

For 2df models, the x1 will not be required. This differs from the evian_logit.

The overdominance model differs from the dominance model in that the D column coded as 0,1,0 for the three genotypes AA,Aa,aa, represents the dominance deviation from additivity, rather than specifying that a particular allele is dominant or recessive. That is, the D term is fitted jointly with the Additive term in a single model. In summary, the overdominance genotypic model involves two variables representing an additive effect and a dominance deviation;

```
  Overdominance model
        A  D
    AA  0  0
    AB  1  1
    BB  2  0
```

In this case, the column of interest is D, and should be inserted in as the X vector. Both userformulaX and userformulaY will need to contain the A column in the right hand side of the formula, and be treated in effect as a continuous covariate. The A column would also need to be present in the Z data frame.

For the alternate genotypic model 2df, the two individual terms G1, G2 use a different version of "genotypic" coding, ie. use of dummy variables to represent genotypes.

```
Genotypic model
     G1 G2
  AA  0  0
  AB  1  0
  BB  0  1
```

In this case, to run the full 2df genotypic model, you will need to run this function twice; once with G1 as the column of interest and another time with G2. Of course, the column that is not of interest would need to go into the right hand side of both `userformulaX and userformulaY` and be treated in effect as a covariate. And again, this column would need to also be present in the `Z` data frame.

These two models are implemented as in plink (AJHG:81; http://pngu.mgh.harvard.edu/purcell/plink/).

Exact variable syntax and character spacing in `formula` is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in `formula`.

Value

This function will produce a plot in a X11 window or a Quartz device window for MacOSX.

Note

When `lolim/hilim` is NOT defined, then the boundaries of the beta grid will be determined by the default `bse=5`, or a `bse` defined by the user. Otherwise, the user can define the exact beta grid boundaries using `lolim/hilim`.

In some cases the beta grid (using `bse` or `lolim/hilim`) may need to be increased substantially (`bse` as large as 15) if covariates are present in the formula.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Blume JD, et. al. Stat Med. 26(15):2919-36.

See Also

`evian_linear`, `evian_linear_plotsnp`, `linear_plotsnp li_plot`

Examples

```
## Not run:
## Computes profile LF for coefficient of specified column (genotypes, xcol) of regressic
linear_plotsnp (X, Y, Z, cluster.id, userformulaX,
            userformulaY, m=1000, bse=5,
            lolim, hilim, robust=FALSE)
## End(Not run)
```

`linear_snp` *Linear regression likelihood function called in evian_linear*

Description

This is the core linear likelihood function that computes the likelihood intervals for each SNP. This function is used for the additive, dominant, recessive, 2df overdominance and 2df genotypic models.

Usage

```
linear_snp(X, Y, Z, cluster.id, userformulaX,
           userformulaY, m=1000, bse=5,
           lolim, hilim, robust=FALSE)
```

Arguments

`X`	a vector of snp genotype data coded, `0,1,2,NA`
`Y`	a vector of the continuous outcome variable
`Z`	a data frame with the first column containing the "X intercept" $vectorof1's$, and followed by columns of x predictor data $covariates$. The header for the first column of intercept must be 'intercept' because `Z$intercept` is used in the variables `userformulaX, userformulaY` .
`cluster.id`	a vector of family ids, that groups a number of related individuals.
`userformulaX`	string containing the glm formula for X, ie. `X ~ Z$intercept + Z$weight` . Use `as.factor(Z$var)` if the data is categorical.
`userformulaY`	string containing the glm formula for Y, ie. `Y ~ -1 + Z$intercept + Z$weight`. Use `as.factor(Z$var)` if the data is categorical.
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e. default = `5`.
`lolim`	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the regression parameter B, at which to calculate the likelihood function.
`hilim`	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the regression parameter B, at which to calculate the likelihood funciton.
`robust`	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. cluster.id, or FID.

Details

This core function is run iteratively in a loop within `evian_linear` for the following selected models: additive, dominant, recessive, 2df overdominance and genotypic.

Depending on which 1df model is desired, the SNP data `X` will need to be coded accordingly. Coding should be as follows:

```
  Additive
AA  0
AB  1
BB  2

  Dominant
AA  0
AB  1
BB  1

  Recessive
AA  0
AB  0
BB  1
```

For 2df models, the `x1` will not be required. This differs from the `evian_logit`.

The overdominance model differs from the dominance model in that the `D` column coded as 0,1,0 for the three genotypes AA,Aa,aa, represents the dominance deviation from additivity, rather than specifying that a particular allele is dominant or recessive. That is, the `D` term is fitted jointly with the Additive term in a single model. In summary, the overdominance genotypic model involves two variables representing an additive effect and a dominance deviation;

```
Overdominance model
      A  D
  AA  0  0
  AB  1  1
  BB  2  0
```

In this case, the column of interest is D, and should be inserted in as the `X` vector. Both `userformulaX and userformulaY` will need to contain the A column in the right hand side of the formula, and be treated in effect as a continuous covariate. The A column would also need to be present in the `Z` data frame.

For the alternate genotypic model 2df, the two individual terms `G1, G2` use a different version of "genotypic" coding, ie. use of dummy variables to represent genotypes.

```
Genotypic model
     G1 G2
  AA  0  0
  AB  1  0
  BB  0  1
```

In this case, to run the full 2df genotypic model, you will need to run this function twice; once with G1 as the column of interest and another time with G2. Of course, the column that is not of interest would need to go into the right hand side of both `userformulaX and userformulaY` and be treated in effect as a covariate. And again, this column would need to also be present in the `Z` data frame.

These two models are implemented as in plink (AJHG:81; http://pngu.mgh.harvard.edu/purcell/plink/).

Exact variable syntax and character spacing in `formula` is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in `formula`.

Value

A vector with eight elements as follows: maxlr, mle, k32_low, k32_hi, k100_low, k100_hi, k1000_low, k1000_hi

The six k_low/hi elements are the limits of the 3 likelihood intervals, ie. 1/k1, 1/k2, 1/k3 intervals.

Note

When `lolim/hilim` is NOT defined, then the boundaries of the beta grid will be determined by the default `bse=5`, or a `bse` defined by the user. Otherwise, the user can define the exact beta grid boundaries using `lolim/hilim`.

In some cases the beta grid (using `bse` or `lolim/hilim`), may need to be increased substantially (`bse` as large as 15) if covariates are present in the formula.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Blume JD, et. al. Stat Med. 26(15):2919-36.

See Also

`evian_linear`, `evian_linear_plotsnp`, `linear_plotsnp` `li_plot`

Examples

```
## Not run:
## Computes profile LF for coefficient of specified column (genotypes, xcol) of regressio
linear_snp (X, Y, Z, cluster.id, userformulaX,
            userformulaY, m=1000, bse=5,
            lolim, hilim, robust=FALSE)
## End(Not run)
```

---

li_plot *Plotting, zooming, and saving likelihood interval plots*

Description

Plots the likelihood intervals for genetic association data in a defined chromosomal region, and/or saves a .png file.

Usage

```
li_plot (bpstart = 0, bpend = 1000000000, dframe, title = "My model",
         showmaxlr = 3, kcutoff = 32, yaxislim = 10, grays = TRUE,
         pl = 'logit', pngfile = FALSE)
```

Arguments

bpstart	numeric, the start base position on the chromosome.
bpend	numeric, the end base position on the chromosome.
dframe	data frame, the output data frame from evian_logit.
title	string, a title for the plot.
showmaxlr	specifies the number of SNPs for which the user wants the maximum likelihood ratios displayed, default = 3.
kcutoff	the strength of evidence criterion k: 32, 100, 1000, default = 32.
yaxislim	y axis limit for the beta coefficient in linear and odds ratio in logistic, default = 10.
grays	logical, if TRUE, then individual intervals for a given SNP will be in gray if the 1/k likelihood interval includes OR = 1 as a plausible value.
pl	string, logit for logistic regression plot or linear for linear regression plot.
pngfile	logical, if TRUE, then the likelihood plot will be saved as .png file in the working directory.

Details

This function is called by evian_logit by default to produce a plot in an X11 window, and/or save it as a .png file if pngfile=TRUE. It allows the user to view pre-defined regions of the full plot provided by evian_logit. This is a zoom feature. The filename will be a time stamp and saved in the current working directory.

Value

This function will produce a plot in a X11 window or a Quartz device window for MacOSX or a .png image file.

In the plot, the SNPs with OR = 1 in the 1/k likelihood interval will be in gray if grays = TRUE. Those with OR >= 1 not in the 1/k likelihood interval will be in color with the 1/32 intervals in green, 1/100 in red and 1/1000 in blue.

Note

You may need to quit R before the image file is written to disk.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Strug et. al, 2009. EJHG, 17:1171-1181;

See Also

evian_logit, evian_logit_plotsnp, logit_snp, logit_plotsnp

Examples

```
## Not run:
li_plot (bpstart = 0, bpend = 1000000000, dframe, title = "My model",
         showmaxlr = 3, kcutoff = 32, yaxislim = 10, grays = TRUE,
         pl = 'logit', pngfile = FALSE)
## End(Not run)
```

`logit_plotsnp` *Logistic regression likelihood plotting function called in evian_logit_plotsnp*

Description

This is the core logistic likelihood function that plots the likelihood function for a single SNP. This function is used for the additive, dominant, recessive, and 2df overdominance and genotypic models.

Usage

```
logit_plotsnp (snp, y, x, x1, Z, userformulaA,
               robust = FALSE, rfdum = 'alphanum',
               m = 200, bse, lolim = log(0.025),
               hilim = log(4), k1 = 32, k2 = 100,
               print.details = TRUE)
```

Arguments

`snp`	string, the snp name of interest stored in the `map` data frame, e.g. rs21547.
`y`	a vector of caseflag status for each sample, `0,1`.
`x`	a vector of snp genotype data coded, `0,1,2,NA`.
`x1`	a vector of snp genotype data coded for genotypic models with two degrees of freedom, `0,1,2,NA`.
`Z`	a data frame with columns of x predictor data *covariates*. The column names of Z are used in defining `userformulaA`.
`userformulaA`	string containing the glm formula for Y, e.g., `Y ~ x + Z$age` for models where the snp covariate is a 1df parameterization, or `Y ~ x + x1 + Z$age` for 2df snp parameterization. Use `as.factor(Z$var)` if the variable is categorical. Note, the formula must be at least `Y ~ x` for 1df, and `Y ~ x + x1` for 2df.
`robust`	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
`rfdum`	string to designate the reference group for categorical variables. Options: `alphanum` orders levels alphanumerically, `most` chooses the level observed most frequently, `least` chooses the level observed least frequently.
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e.

`lolim` numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the log(OR), at which to calculate the likelihood function.

`hilim` numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the log(OR), at which to calculate the likelihood funciton.

`k1` numeric, 1/k1 likelihood interval: `8,32,100,1000`, plotted in green

`k2` numeric, 1/k2 likelihood interval: `8,32,100,1000`, plotted in red

`print.details` logical, if `TRUE`, then displays plot legend and details

Details

This core function is run within `evian_logit_plotsnp` for the following selected models: additive, dominant, recessive, 2df overdominance and genotypic.

Depending on which 1df model is desired, the SNP data `x` will need to be coded accordingly. The `x1` variable will be ignored. Coding should be as follows:

```
  Additive
AA  0
AB  1
BB  2

  Dominant
AA  0
AB  1
BB  1

  Recessive
AA  0
AB  0
BB  1
```

For 2df models, both `x` and `x1` variables will be required. The overdominance model differs from the dominance model in that the `x` variable (variable of interest) coded as 0, 1, 0 for the three genotypes AA, Aa, aa, represents the dominance deviation from additivity, rather than specifying that a particular allele is dominant or recessive. That is, the `x` term is fitted jointly with the Additive term in a single model. In summary, the overdominance genotypic model involves two variables representing an additive effect and a dominance deviation;

```
Overdominance model
      A(x1) D(x)
  AA  0     0
  Aa  1     1
  aa  2     0
```

For the alternate 2df genotypic model, the two individual terms `(x1 - G1, x2 - G2` use a different version of "genotypic" coding, ie.i use of dummy variables to represent genotypes.

```
Genotypic model
      G1(x1) G2(x2)
  AA   0      0
  Aa   1      0
  aa   0      1
```

These two models are implemented as in plink (AJHG:81; http://pngu.mgh.harvard.edu/purcell/plink/).

The `formula` must be at least `Y ~ x` for 1df models, and `Y ~ x + x1` for 2df models. Exact variable syntax and character spacing in `formula` is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in `formula`.

Value

This function will produce a plot in a X11 window or a Quartz device window for MacOSX.

Note

When the number of beta standard errors, `bse` is NOT defined, then the boundaries of the beta grid will use the default `lolim/hilim` values, or those defined by the user. However, if `bse` is defined, then `lolim/hilim` values will be ignored.

The beta grid limits may need to be adjusted or broadened in plots when the calculated likelihood intervals are not available, e.g., NAs provided.

In some cases the beta grid (using `bse` or `lolim/hilim`), may need to be increased substantially (`bse` as large as 15) if covariates are present in the formula.

For a given plot, the typical runtime for 1df models, ie. additive, dominant or recessive, without any covariates on a single processor of a 2.53 GHz Intel Core 2 Duo will be around 5 seconds for a dataset with 250 individuals.

Finally, estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Enable the X11 server on your terminal before running R. If your desktop OS is Linux or MacOSX (10.5), the X11 server should be automatic. In MS-Windows, you may need launch an X server appliction such as Xming.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Royall, R.M,. Statistical evidence: a likelihood paradigm. ISBN:9780412044113. Edwards, A.W.F., Likelihood ISBN:0801844436. Strug & Hodge, I. Human Heredity, 61:166-188; Strug & Hodge, II. Human Heredity, 61:200-209;

See Also

`evian_logit`, `evian_logit_plotsnp`, `logit_snp`, `li_plot`

Examples

```
## Not run:
logit_plotsnp (snp, y, x, x1, Z, userformulaA, rfdum = 'alphanum',
               m = 200, n = 1, bse, lolim = log(0.025), hilim = log(4),
               k1 = 32, k2 = 100, print.details = TRUE)

## End(Not run)
```

`logit_snp` *Logistic regression likelihood function called in evian_logit*

Description

This is the core logistic likelihood function that computes the likelihood intervals for each SNP. This function is used for the additive, dominant, recessive, 2df overdominance and 2df genotypic models.

Usage

```
logit_snp (y, x, x1, Z, userformulaA,
           robust = FALSE, rfdum = 'alphanum',
           m = 200, bse, lolim = log(0.025),
           hilim = log(4), k1 = 32, k2 = 100,
           k3 = 1000)
```

Arguments

`y`	a vector of caseflag status for each sample, `0,1`.
`x`	a vector of snp genotype data coded, `0,1,2,NA`
`x1`	a vector of snp genotype data coded for genotypic models with two degrees of freedom, `0,1,2,NA`
`Z`	a data frame with columns of x predictor data *covariates*. The column names of Z are used in defining `userformulaA`
`userformulaA`	string containing the glm formula for Y, e.g., `Y ~ x + Z$age` for models where the snp covariate is a 1df parameterization, or `Y ~ x + x1 + Z$age` for 2df snp parameterization. Use `as.factor(Z$var)` if the variable is categorical. Note, the formula must be at least `Y ~ x` for 1df, and `Y ~ x + x1` for 2df.
`robust`	logical, if `TRUE`, then a robust adjustment is applied to the likelihood function to account for the cluster nature in the data, e.g. family id, FID.
`rfdum`	string to designate the reference group for categorical variables. Options: `alphanum` orders levels alphanumerically, `most` chooses the level observed most frequently, `least` chooses the level observed least frequently.
`m`	numeric, the density of the grid at which to compute the standardized likelihood function. A beta grid is defined as the grid of values for the snp parameter at which to evaluate the likelihood function.
`bse`	numeric, the number of beta standard errors to utilize in constraining the beta grid limits, e.g., beta grid is evaluated at $\hat{\beta}$ +/- 5 s.e.
`lolim`	numeric, instead of the number of s.e., specify the lower limit for the grid, or the minimum value of the log(OR), at which to calculate the likelihood function.
`hilim`	numeric, instead of the number of s.e., specify the upper limit for the grid, or the maximum value of the log(OR), at which to calculate the likelihood funciton.
`k1`	numeric, 1/k1 likelihood interval: `8,32,100,1000`, plotted in green
`k2`	numeric, 1/k2 likelihood interval: `8,32,100,1000`, plotted in red
`k3`	numeric, 1/k3 likelihood interval: `8,32,100,1000`, plotted in blue

Details

This core function is run iteratively in a loop within `evian_logit` for the following selected models: additive, dominant, recessive, 2df overdominance and genotypic.

Depending on which 1df model is desired, the SNP data `x` will need to be coded accordingly. The `x1` variable will be ignored. Coding should be as follows:

```
  Additive
AA  0
AB  1
BB  2

  Dominant
AA  0
AB  1
BB  1

  Recessive
AA  0
AB  0
BB  1
```

For 2df models, both `x` and `x1` variables will be required. The overdominance model differs from the dominance model in that the `x` variable (variable of interest) coded as 0, 1, 0 for the three genotypes AA, Aa, aa, represents the dominance deviation from additivity, rather than specifying that a particular allele is dominant or recessive. That is, the `x` term is fitted jointly with the Additive term in a single model. In summary, the overdominance genotypic model involves two variables representing an additive effect and a dominance deviation;

```
Overdominance model
      A(x1) D(x)
  AA  0     0
  Aa  1     1
  aa  2     0
```

For the alternate 2df genotypic model, the two individual terms `x1 - G1, x2 - G2` use a different version of "genotypic" coding, ie.use of dummy variables to represent genotypes.

```
Genotypic model
      G1(x1) G2(x2)
  AA   0      0
  Aa   1      0
  aa   0      1
```

These two models are implemented as in plink (AJHG:81; http://pngu.mgh.harvard.edu/purcell/plink/).

The `formula` must be at least `Y ~ x` for 1df models, and `Y ~ x + x1` for 2df models. Exact variable syntax and character spacing in `formula` is important for correct parsing and computation. In other words, a single whitespace separates each variable or symbol in `formula`.

Value

A vector with nine elements as follows:

maxlr, mle, k32_low, k32_hi, k100_low, k100_hi, k1000_low, k1000_hi, flip

The six k_low/hi elements are the limits of the 3 likelihood intervals, ie. 1/k1, 1/k2, 1/k3 intervals.

When the value of mle is less than one, the k intervals are inverted and flip=1, otherwise flip=0.

Note

When the number of beta standard errors `bse` is NOT defined, then the boundaries of the beta grid will use the default `lolim/hilim` values, or those defined by the user. However, if `bse` is defined, then `lolim/hilim` values will be ignored.

In some cases the beta grid (using `bse` or `lolim/hilim`), may need to be increased substantially (`bse` as large as 15) if covariates are present in the formula.

Estimation may become inaccurate with large number of correlated covariates, similar to known limitation of profile likelihoods.

Author(s)

Dr. Lisa J Strug ⟨lisa.strug@utoronto.ca⟩

References

Strug, Rohde & Corey. Am Stat, 61:207-212; Strug & Hodge, I. Human Heredity, 61:166-188; Strug & Hodge, II. Human Heredity, 61:200-209;

See Also

`evian_logit`, `evian_logit_plotsnp`, `logit_plotsnp`, `li_plot`

Examples

```
## Not run:
logit_snp (y, x, x1, Z, userformulaA, rfdum = 'alphanum',
           m = 200, n = 1, bse, lolim = log(0.025),
           hilim = log(4), k1 = 32, k2 = 100, k3 = 1000)

## End(Not run)
```

`robust_for_cluster` *Robust adjustment function called in logit_snp*

Description

The robust function computes an adjustment that is applied to the likelihood function to account for the cluster nature in the data.

Usage

```
robust_for_cluster (x, y, Z, beta)
```

Arguments

x	a vector of snp genotype data coded, 0,1,2,NA
y	a vector of caseflag status for each sample, 0,1.
Z	a data frame with columns of x predictor data *covariates* or dummy variables if the covariate is categorical. The first column of Z should be the family ID column, e.g. FID. The subsequent column names of Z should appear in the same order as the covariates in userformulaA.
beta	a vector of beta coefficients. Usually the coefficient output from the glm fit of userformulaA.

Details

The robust function is called from within evian_logit and logit_snp functions. It computes an adjustment that is applied to the likelihood function to account for the cluster nature in the data. Therefore it is important to have the family ID column, e.g. FID.

Value

A numeric constant.

Note

Ensure that x, y, and Z have identical lengths. Often missing values coded as NA may cause problems. If this function is called manually, then it may be worthwhile to run na.omit first.

Author(s)

Zeynep Baskurt ⟨zeynep@utstat.utoronto.ca⟩

References

Strug & Hodge I & II; Strug, Rohde & Corey.

See Also

evian_logit, evian_logit_plotsnp, logit_plotsnp, li_plot

Examples

```
## Not run:
robust_for_cluster (x, y, Z, beta)

## End(Not run)
```

Index