Package 'ehpoly'

January 2, 2017

Type Package

Title Test for Etiologic Heterogeneity

Version 0.1.1

Description Use polytomous logistic regression to test for etiologic heterogeneity in risk factors both across disease subtypes and across individual tumor markers.

Depends R (>= 3.1.0)

License GPL-2

Encoding UTF-8

LazyData true

Imports nnet, aod

RoxygenNote 5.0.1

Author Emily Zabor [aut, cre]

Maintainer Emily Zabor <zabore@mskcc.org>

R topics documented:

---

ehpoly *Conduct an analysis of etiologic heterogeneity using polytomous logistic regression*

Description

ehpoly takes a list of individual tumor markers and a list of risk factors and returns results related to the question of whether each risk factor differs across levels of the disease subtypes and the question of whether each risk factor differs across levels of each individual tumor marker of which the disease subtypes are comprised.

Input is a dataframe that contains the individual tumor markers and the risk factors of interest. The tumor markers must be binary and must have levels 0 or 1 for cases. The tumor markers must have a value of 999 for control subjects. For categorical tumor markers, a reference level should be selected and then indicator variables for each remaining level of the tumor marker should be created. For continuous tumor markers, categories should be formed and then indicator variables

can be constructed as in the case of categorical tumor markers. Risk factors can be either binary or continuous. For categorical risk factors, a reference level should be selected and then indicator variables for each remaining level of the risk factor should be created. Categorical risk factors entered as is will be treated as ordinal.

Usage

```
ehpoly(tm, rf, case, df)
```

Arguments

tm	a list of the names of the binary tumor markers. Each must have levels 0 or 1 for case subjects. This value will be missing for all control subjects.
rf	a list of the names of the binary or continuous risk factors. For binary risk factors the lowest level will be used as the reference level.
case	denotes the variable that contains each subject's status as a case or control. This value should be 1 for cases and 0 for controls. Argument must be supplied in quotes.
df	the name of the dataframe that contains the tumor markers and risk factors.

Value

Returns a list.

beta is a matrix containing the estimated beta parameters with a column for each risk factor and a row for each disease subtype.

beta_se contains the associated standard errors.

eh_pval is a vector of p-values for testing whether each risk factor differs across levels of the disease subtype.

gamma is a matrix containing the estimated gamma parameters, obtained as linear combinations of the beta parameters with a column for each risk factor and a row for each tumor marker.

gamma_se contains the associated standard errors.

gamma_p is a matrix of p-values for testing whether each risk factor differs across levels of each tumor marker, with a column for each risk factor and a row for each tumor marker.

or_ci_p is a dataframe with a odds ratio (95 factor/subtype combination, as well as a column of etiologic heterogeneity p-values.

beta_se_p is a dataframe with the estimated beta parameters (SE) for each risk factor/subtype combination, as well as a column of etiologic heterogeneity p-values.

gamma_se_p is a dataframe with estimates of the gamma tumor marker effects (SE) and their associated p-values.

Author(s)

Emily C Zabor <zabore@mskcc.org>

Index