Package ‘mztwinreg’

February 20, 2015

Title Regression Models for Monozygotic Twin Data

Description Linear and logistic regression models for quantitative genetic analysis of data from monozygotic twins.

Version 1.0-1

Date 2015-01-13

Author Aldo Cordova-Palomera

Maintainer Aldo Cordova-Palomera <`a.cordova@ub.edu`>

Imports rms, mclogit

Suggests lme4

License GPL-3

URL `https://github.com/AldoCP/mztwinreg`

NeedsCompilation no

Repository CRAN

Date/Publication 2015-01-14 19:50:57

R topics documented:

`mztwinreg-package` *Regression Modeling of MZ Twin Data*

Description

Linear and logistic regression models for quantitative genetic analysis of data from monozygotic twins

Details

Package:	mztwinreg
Type:	Package
Version:	1.0
Date:	2015-01-13
License:	What license is it under?

This package implements a number of functions to perform different linear and logistic regression models of data from monozygotic twins. Examples include the variability gene approach by Berg, K. (1994), the Epigenetic Epidemilogy model by Tan, Q. (2013)., and the conventional approaches of Carlin, J. B. et al. (1994).

Author(s)

Aldo Cordova-Palomera

Maintainer: Aldo Cordova-Palomera <a.cordova@ub.edu>

References

Berg, K. (1994). Gene-environment interaction: variability gene concept. In Genetic factors in coronary heart disease (pp. 373-383). Springer Netherlands.

Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. International Journal of Epidemiology, 34(5), 1089-1099.

Cordova-Palomera, A., Alemany, S., Fatjo-Vilas, M., Goldberg, X., Leza, J. C., Gonzalez-Pinto, A., Nenadic, I., & Fananas, L. (2014). Birth Weight, Working Memory and Epigenetic Signatures in IGF2 and Related Genes: A MZ Twin Study. PloS one, 9(8), e103639.

Cordova-Palomera, A., Fatjo-Vilas, M., Kebir, O., Gasto, C., Krebs, M. O., & Fananas, L. (2014). Polymorphic variation in the epigenetic gene DNMT3B modulates the environmental impact on cognitive ability: A twin study. European Psychiatry.

Tan, Q. (2013). Epigenetic Epidemiology of Complex Diseases Using Twins. Medical Epigenetics, 1(1), 46-51.

Tan, Q., Frost, M., Heijmans, B. T., von Bornemann Hjelmborg, J., Tobi, E. W., Christensen, K., & Christiansen, L. (2014). Epigenetic signature of birth weight discordance in adult twins. BMC genomics, 15(1), 1062.

`abs_dif` *Test for intrapair phenotypic differences based on pair-level predictor variables*

Description

Regression models to test whether the intrapair phenotypic differences are due to factors that are identical for both co-twins (i.e., a genotype). It is based on the work on "variability genes", by Berg, K. (1994).

Usage

```
abs_dif(formula, regression = "linear", data, ...)
```

Arguments

`formula`	an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. See `formula`.
`regression`	the type of regression model to be fitted. Use either `'linear'` (default) or `'logistic'`.
`data`	a data frame containing the variables in the model.
`...`	additional arguments to be passed to either `ols` (linear regression) or `lrm` (logistic model), from the `rms` package.

Details

Inspired by the "variability gene" concept by Berg, K. (1994), these regression models (linear and logistic) allow determining whether a pair-level variable (i.e., a genotype, which is identical for both co-twins). Further usage details are explained by Cordova-Palomera, A. et al. (2014).

Value

`abs_dif` returns an object of `class` "rms", along with either c("ols", "lm") (linear models) or c("lrm", "glm") (logistic models).

Author(s)

Inspired by Berg (1994). Developed by Aldo Cordova-Palomera. See Cordova-Palomera et al. (2014).

References

Berg, K. (1994). Gene-environment interaction: variability gene concept. In Genetic factors in coronary heart disease (pp. 373-383). Springer Netherlands.

Cordova-Palomera, A., Fatjo-Vilas, M., Kebir, O., Gasto, C., Krebs, M. O., & Fananas, L. (2014). Polymorphic variation in the epigenetic gene DNMT3B modulates the environmental impact on cognitive ability: A twin study. European Psychiatry.

See Also

`rms`, `ols`, `lrm`

Examples

```
data(flu_weight)

# The linear regression below tests whether the intrapair differences in
# DNA methylation (outcome) depend on pair-specific variables such as gender,
# age or both.
# The conceptual justification of the working hypothesis is partly inspired
# by Fraga, M. F., et al. (2005) Proceedings of the National Academy of
# Sciences of the United States of America, 102(30), 10604-10609.
(variability_linear <- abs_dif(DNAmeth ~ Gender + Age, data=flu_weight, regression='linear'))

# The logistic regression below tests whether or not there was an
# increased/decreased rate of phenotypic differences in flu liability during
# childhood in males, compared with females.
# Namely, it evaluates there were more intrapair differences in flu liability
# in one of the genders (male or female).
(variability_logistic <- abs_dif(FluChild ~ Gender, data=flu_weight, regression='logistic'))
```

`env_dif` *Test for intrapair phenotypic differences in an outcome based on intrapair-differences predictor variables*

Description

Linear regression model to test whether the intrapair differences in an outcome phenotype are due to intrapair differences in a predictor phenotype.

Usage

```
env_dif(formula, data, ...)
```

Arguments

`formula`	an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. See `formula`.
`data`	a data frame containing the variables in the model.
`...`	additional arguments to be passed to `ols`, from the `rms` package.

Details

Inspired by the implementation of Carlin, J. B. et al. (1994), this linear regression model without intercept tests allows evaluating whether intrapair differences in one or more predictor variables are related to intrapair differences in the outcome.

Value

`env_dif` returns an object of `class` "lm".

Author(s)

Developed by Aldo Cordova-Palomera, following Carlin, J. B. et al. (1994).

References

Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. International Journal of Epidemiology, 34(5), 1089-1099.

See Also

`lm`

Examples

```
data(flu_weight)

# The linear regression below tests whether the intrapair differences in weight
# (in kilograms) of a given twin-pair predict its intrapair differences in DNA
# methylation levels at a given genomic locus.
summary(unique_env_linear <- env_dif(DNAmeth ~ Weight, data=flu_weight))
```

`env_dif_logistic` *Test for intrapair phenotypic differences in a binary outcome based on intrapair-differences predictor variables*

Description

Conditional logit model to test whether the intrapair differences in an outcome phenotype (binary variable) are due to intrapair differences in a predictor phenotype.

Usage

```
env_dif_logistic(formula, cluster = "default", data, ...)
```

Arguments

`formula`	an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. See `formula`.
`cluster`	a vector identifying the pairs in the sample. If not specified, the default value assumes that twin pairs are introduced in adjacent rows.
`data`	a data frame containing the variables in the model.
`...`	additional arguments to be passed to `mclogit`, from the `mclogit` package.

Details

This is an R implementation of some code proposed by Carlin, J. B. et al. (1994) for Stata.

Note that this conditional logit model -either in the original version by Carlin, J. B. et al. (1994) or in the current R scripts- is not widely used and may require improvements.

Value

`env_dif_logistic` returns an object of `class` c("mclogit", "lm").

Author(s)

Carlin, J. B. et al. (1994) proposed this model and implemented it in Stata. Aldo Cordova-Palomera adjusted the algorithm for R.

References

Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. International Journal of Epidemiology, 34(5), 1089-1099.

See Also

`mclogit`, `lm`, `lrm`

Examples

```
data(flu_weight)

# The linear regression below tests whether the intrapair differences in flu
# presence (binary outcome) of a given twin-pair are predicted by its
# differences in weight (in kilograms).
summary(unique_env_logistic <- env_dif_logistic(FluNow ~ Weight, data=flu_weight))
```

`fam_env` *Linear and logistic regression models for familiar and unique environmental factors*

Description

This function implements linear and logistic regression models to test for the association between an outcome phenotype and both the familial (genes + shared environment) and the unique environmental influences on a predictor variable.

Usage

```
fam_env(formula, BbBw = NULL, regression = "linear", cluster = "default",
        adjust = "robcov", robcov_method = "huber", bootcov_B = 200, data, ...)
```

Arguments

`formula`	an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. See `formula`.
`BbBw`	the name of the predictor variable(s) (already in `formula`) for which the familial and the unique environmental influences are going to be evaluated.
`regression`	the type of regression model to be fitted. Use either `'linear'` (default) or `'logistic'`.
`cluster`	a vector identifying the pairs in the sample. If not specified, the default value assumes that twin pairs are introduced in adjacent rows.
`adjust`	a method to adjust for correlated responses (heteroskedasticity) of twin pairs. Use either `'robcov'` (default) or `'bootcov'`.
`robcov_method`	if `adjust = 'robcov'` (default) is selected, it allows choosing a method to adjust the variance-covariance matrix. Use either `'huber'` (default: Huber-White sandwich estimator) or `'efron'` (especially for small samples; see `robcov`).
`bootcov_B`	if `adjust = 'bootcov'` is selected, it allows specifying the number of bootstrap repetitions to computes an estimate of the covariance matrix for a set of regression coefficients.
`data`	a data frame containing the variables in the model.
`...`	additional arguments to be passed to either `ols` (linear regression) or `lrm` (logistic model), from the `rms` package.

Details

As shown by Carlin, J. B. et al. (1994), these cluster-based regression models allow parsing out familial and environmental factors contributing to the value of a predictor variable.

Value

`fam_env` returns an object of `class` "rms", along with either c("ols", "lm") (linear models) or c("lrm", "glm") (logistic models).

Author(s)

Developed by Aldo Cordova-Palomera, following Carlin, J. B. et al. (1994).

References

Carlin, J. B., Gurrin, L. C., Sterne, J. A., Morley, R., & Dwyer, T. (2005). Regression models for twin studies: a critical review. International Journal of Epidemiology, 34(5), 1089-1099.

See Also

`rms`, `ols`, `lrm`, `robcov`, `bootcov`

Examples

```
data(flu_weight)

# The linear regression below tests whether the intrapair differences in
# DNA methylation of a given twin-pair are predicted by its differences in
# either familial (Bb) or environmental (Bw) factors influencing weight
# (in kilograms).
(fam_env_linear <- fam_env(DNAmeth ~ Gender + Age + Weight, BbBw="Weight",
regression='linear', data=flu_weight))

# The linear regression below tests whether the intrapair differences in
# flu (binary outcome) of a given twin-pair are predicted by its differences
# in either familial (Bb) or environmental (Bw) factors influencing weight
# (in kilograms).
(fam_env_logistic <- fam_env(FluNow ~ Gender + Age + Weight, BbBw="Weight",
regression='logistic', data=flu_weight))
```

`flu_weight` *Flu, Body Weight and DNA methylation*

Description

An artificial dataset containing information on flu, body weight and DNA methylation at an immune system gene locus, as well as demographic variables.

Usage

```
data("flu_weight")
```

Format

A data frame with 200 observations on the following 7 variables.

`DNAmeth` a numeric vector: DNA methylation fraction (ranging 0-1) at an immune system gene locus

`FluNow` a numeric vector: Presence of severe flu at the moment of sample collection

`FluChild` a numeric vector: History of recurrent flu during childhood

`Gender` a character vector: Pair's gender

`Age` a numeric vector: Pair's age

`Weight` a numeric vector: Body weight in kilograms

`PairNum` a numeric vector: Randomly assigned pair number

Details

As observed from this dataset, the individuals of the sample are sorted following two criteria: 1) twin pairs are put together (adjacent rows), and 2) all twins are discordant for present flu (`FluNow`).

The second condition is necessary only when using the `log_dif` function.

It is recommended pre-sorting all datasets used along with this R package following criterion "1" or criteria "1" and "2".

`log_dif` *Test for intrapair logarithm-based phenotypic differences based on pair-level predictor variables*

Description

Linear regression model to test whether the intrapair phenotypic differences in a continuous outcome (i.e., DNA methylation) in phenotype-discordant pairs are due to factors that are identical for both co-twins (i.e., a genotype). It is based on the work on Epigenetic Epidemiology by Tan, Q. (2013). It is similar to the `abs_dif` function in this package, inspired by Berg, K. (1994).

Usage

```
log_dif(formula, data, ...)
```

Arguments

`formula`	an object of class `formula` (or one that can be coerced to that class): a symbolic description of the model to be fitted. See `formula`.
`data`	a data frame containing the variables in the model.
`...`	additional arguments to be passed to either `ols`, from the `rms` package.

Details

Similar to the model implemented in the `abs_dif` function in this package, `log_dif` tests whether the logarithm of the difference in an outcome measure in phenotype-discordant pairs is explained by paired-level exposures such as age. Of note, the affected co-twin is set before (one row above of) his/her co-twin in the `data` dataset. This model was originally proposed by Tan, Q. (2013) in the context of Epigenetic Epidemiology, to evaluate whether phenotype-discordant pairs have DNA methylation differences due to pair-level exposures.

Value

`log_dif` returns an object of `class` c("ols" "rms" "lm").

Author(s)

Aldo Cordova-Palomera.

References

Tan, Q. (2013). Epigenetic Epidemiology of Complex Diseases Using Twins. Medical Epigenetics, 1(1), 46-51.

Tan, Q., Frost, M., Heijmans, B. T., von Bornemann Hjelmborg, J., Tobi, E. W., Christensen, K., & Christiansen, L. (2014). Epigenetic signature of birth weight discordance in adult twins. BMC genomics, 15(1), 1062.

See Also

`rms`, `ols`

Examples

```
data(flu_weight)

# The linear regression below tests whether DNA methylation differences at a
# given locus are predicted by pair level variables (such as gender, age
# or both).
(logarithm_differences <- log_dif(DNAmeth ~ Gender + Age, data=flu_weight))
```

Index