

Package ‘gWQS’

May 7, 2018

Type Package

Title Generalized Weighted Quantile Sum Regression

Version 1.1.0

Author Stefano Renzetti, Paul Curtin, Allan C Just, Ghalib Bello, Chris Gennings

Maintainer Stefano Renzetti <stefano.renzetti88@gmail.com>

Description Fits Weighted Quantile Sum (WQS) regressions for continuous or binomial outcomes.

Imports Rsolnp, ggplot2, ztable, tableHTML

License GPL (>= 2)

LazyData true

RoxygenNote 6.0.1

Suggests knitr, rmarkdown, pander

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2018-05-07 16:38:54 UTC

R topics documented:

gwqs	2
wqs_data	5
Index	6

Description

Fits Weighted Quantile Sum (WQS) regressions for continuous or binomial outcomes.

Usage

```
gwqs(formula, mix_name, data, q = 4, validation = 0.6, valid_var = NULL,
      b = 100, b1_pos = TRUE, b1_constr = FALSE, family = "gaussian",
      seed = NULL, wqs2 = FALSE, plots = FALSE, tables = FALSE)
```

Arguments

formula	An object of class formula specifying the relationship to be tested. If no covariates are being tested specify $y \sim \text{NULL}$.
mix_name	A character vector listing the variables contributing to a mixture effect.
data	The data.frame containing the variables to be included in the model.
q	An integer to specify how mixture variables will be ranked, e.g. in quartiles ($q = 4$), deciles ($q = 10$), or percentiles ($q = 100$). If $q = \text{NULL}$ then the values of the mixture variables are taken (these must be standardized).
validation	Percentage of the dataset to be used to validate the model. If $\text{validation} = 0$ then the test dataset is used as validation dataset too.
valid_var	A character value containing the name of the variable that identifies the validation and the training dataset. You previously need to create a variable in the dataset which is equal to 1 for the observations you want to include in the validation dataset and equal to 0 for the observation you want to include in the training dataset. Assign $\text{valid_var} = \text{NULL}$ if you want to let the function create the validation and training dataset by itself.
b	Number of bootstrap samples used in parameter estimation.
b1_pos	A logical value that determines whether weights are derived from models where the beta values were positive or negative.
b1_constr	A logical value that determines whether to apply positive (if $\text{b1_pos} = \text{TRUE}$) or negative (if $\text{b1_pos} = \text{FALSE}$) constraints in the optimization function for the weight estimation.
family	A character value, if equal to "gaussian" a linear model is implemented, if equal to "binomial" a logistic model is implemented.
seed	An integer value to fix the seed, if it is equal to NULL no seed is chosen.
wqs2	A logical value indicating whether a quadratic term should be included in the model ($\text{wqs2} = \text{TRUE}$) or not ($\text{wqs2} = \text{FALSE}$).
plots	A logical value indicating whether plots should be generated with the output ($\text{plots} = \text{TRUE}$) or not ($\text{plots} = \text{FALSE}$).

`tables` A logical value indicating whether tables should be generated in the directory with the output (`tables = TRUE`) or not (`tables = FALSE`). A preview of the estimates of the final weights is generated in the Viewer Pane

Details

gwqs uses the `glm2` function in the **glm2** package to fit the model. The `glm2` package is a modified version of the `glm` function provided and documented in the `stats` package.

The `solnp` optimization function is used to estimate the weights in each bootstrap sample.

The `seed` argument specifies a fixed seed through the `set.seed` function.

The `wqs2` argument includes a quadratic mixture effect in the linear model. In order to test the significance of this term an Analysis of Variance is executed through the `anova` function.

The `plots` argument produces two figures through the `ggplot` function.

Value

gwqs return the results of the WQS regression as well as many other objects and datasets.

<code>fit</code>	A <code>glm2</code> object that summarizes the output of the WQS model, reflecting either a linear or logistic regression depending on how the <code>family</code> parameter was specified (respectively, "gaussian" or "binomial"). The <code>summary</code> function can be used to call and print fit data.
<code>conv</code>	Indicates whether the solver has converged (0) or not (1 or 2).
<code>wb1pm</code>	Matrix of estimated weights, mixture effect parameter estimates and the associated p-values estimated for each bootstrap iteration.
<code>y_adj</code>	Vector containing the y values (dependent variable) adjusted for the residuals of a fitted model adjusted for covariates.
<code>wqs</code>	Vector containing the wqs index for each subject.
<code>index_b</code>	List of vectors containing the rownames of the subjects included in each bootstrap dataset.
<code>data_t</code>	<code>data.frame</code> containing the subjects used to estimate the weights in each bootstrap.
<code>data_v</code>	<code>data.frame</code> containing the subjects used to estimate the parameters of the final model.
<code>final_weights</code>	<code>data.frame</code> containing the final weights associated to each chemical.
<code>fit_2</code>	It is the same as <code>fit</code> , but it contains the results of the regression with the wqs quadratic term. If <code>wqs2 = FALSE</code> , <code>NULL</code> is returned.
<code>aov</code>	Analysis of variance table to test the significance of the wqs quadratic term in the model. If <code>wqs2 = FALSE</code> , <code>NULL</code> is returned.

Author(s)

Stefano Renzetti, Paul Curtin, Allan C Just, Ghalib Bello, Chris Gennings

References

Carrico C, Gennings C, Wheeler D, Factor-Litvak P. Characterization of a weighted quantile sum regression for highly correlated data in a risk analysis setting. *J Biol Agricul Environ Stat.* 2014:1-21. ISSN: 1085-7117. DOI: 10.1007/s13253-014-0180-3. <http://dx.doi.org/10.1007/s13253-014-0180-3>.

Czarnota J, Gennings C, Colt JS, De Roos AJ, Cerhan JR, Severson RK, Hartge P, Ward MH, Wheeler D. 2015. Analysis of environmental chemical mixtures and non-Hodgkin lymphoma risk in the NCI-SEER NHL study. *Environmental Health Perspectives*, DOI:10.1289/ehp.1408630.

Czarnota J, Gennings C, Wheeler D. 2015. Assessment of weighted quantile sum regression for modeling chemical mixtures and cancer risk. *Cancer Informatics*, 2015:14(S2) 159-171 DOI: 10.4137/CIN.S17295.

Examples

```
# we save the names of the mixture variables in the variable "toxic_chems"
toxic_chems = c("log_LBX074LA", "log_LBX099LA", "log_LBX105LA", "log_LBX118LA",
"log_LBX138LA", "log_LBX153LA", "log_LBX156LA", "log_LBX157LA", "log_LBX167LA",
"log_LBX170LA", "log_LBX180LA", "log_LBX187LA", "log_LBX189LA", "log_LBX194LA",
"log_LBX196LA", "log_LBX199LA", "log_LBXD01LA", "log_LBXD02LA", "log_LBXD03LA",
"log_LBXD04LA", "log_LBXD05LA", "log_LBXD07LA", "log_LBXF01LA", "log_LBXF02LA",
"log_LBXF03LA", "log_LBXF04LA", "log_LBXF05LA", "log_LBXF06LA", "log_LBXF07LA",
"log_LBXF08LA", "log_LBXF09LA", "log_LBXPCBLA", "log_LBXTCDLA", "log_LBXHXCLA")

# To run a linear model and save the results in the variable "results". This linear model
# (family="Gaussian") will rank/standardize variables in quartiles (q = 4), perform a
# 40/60 split of the data for training/validation (validation = 0.6), and estimate weights
# over 5 bootstrap samples (b = 3). Weights will be derived from mixture effect
# parameters that are positive (b1_pos = TRUE). A unique seed was specified (seed = 2016) so
# this model will be reproducible, and plots describing the variable weights and linear
# relationship will be generated as output (plots = TRUE). In the end tables describing the
# weights values and the model parameters with the respectively statistics are generated in
# the viewer window
results = gwqs(y ~ NULL, mix_name = toxic_chems, data = wqs_data, q = 4, validation = 0.6,
              b = 3, b1_pos = TRUE, b1_constr = FALSE, family = "gaussian", seed = 2016,
              wqs2 = FALSE, plots = TRUE, tables = TRUE)

# to test the significance of the covariates
summary(results$fit)
```

`wqs_data`*Exposure concentrations of 34 PCB (simulated dataset)*

Description

We created the 'wqs_data' dataset to show how to use this function. These data reflect 34 exposure concentrations simulated from a distribution of PCB exposures measured in subjects participating in the NHANES study (2001-2002). Additionally, an end-point measure, simulated from a distribution of leukocyte telomere length (LTL), a biomarker of chronic disease, is provided as well (variable name: `y`), as well as simulated covariates, e.g. `sex`, and a dichotomous outcome variable (variable name: `disease_state`). This dataset can thus be used to test the 'gWQS' package by analyzing the mixed effect of the 34 simulated PCBs on the continuous or binary outcomes, with adjustments for covariates.

Usage

`wqs_data`

Format

A data frame with 500 rows and 37 variables

Details

`y` continuous outcome, biomarker of chronic disease

disease_state dichotomous outcome, state of disease

sex covariate, gender of the subject

log_LBX 34 exposure concentrations of PCB exposures ...

Index

*Topic **datasets**

wqs_data, [5](#)

anova, [3](#)

ggplot, [3](#)

glm, [3](#)

gwqs, [2](#)

set.seed, [3](#)

solnp, [3](#)

wqs_data, [5](#)