Package 'gwbr'

June 27, 2023

Title Local and Global Beta Regression

Version 1.0.5

Description Fit a regression model for when the response variable is presented as a ratio or proportion. This adjustment can occur globally, with the same estimate for the entire study space, or locally, where a beta regression model is fitted for each region, considering only influential locations for that area. Da Silva, A. R. and Lima, A. O. (2017) <doi:10.1016/j.spasta.2017.07.011>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

NeedsCompilation no

Author Roberto Marques [aut, cre],
Alan da Silva [aut]

Maintainer Roberto Marques <robertomarques_23@yahoo.com.br>

Depends R (>= 3.5.0)

Repository CRAN

Date/Publication 2023-06-27 02:10:02 UTC

R topics documented:

---

betareg_gwbr   *Global Beta Regression Model*

Description

Fits a global regression model using the beta distribution, recommended for rates and proportions, via maximum likelihood using a parametrization with mean (transformed by the link function) and precision parameter (called phi). For more details see Ferrari and Cribari-Neto (2004).

Usage

```
betareg_gwbr(
  yvar,
  xvar,
  data,
  link = c("logit", "probit", "loglog", "cloglog"),
  maxint = 100
)
```

Arguments

yvar	A vector with the response variable name.
xvar	A vector with descriptive variable(s) name(s).
data	A data set object with yvar and xvar.
link	The link function used in modeling. The options are: `"logit"`, `"probit"`, `"loglog"` or `"cloglog"`. The default is `"logit"`.
maxint	A Maximum number of iterations to numerically maximize the log-likelihood function in search of the estimators. The default is `maxint=100`.

Value

A list that contains:

- `parameter_estimates` - Parameter estimates.
- `phi` - Precision parameter estimate.
- `residuals` - Table with observed values (y), estimated values in classical regression (yhatcl), pure residual in classical regression (ecl), estimated values (yhat), the link function applied in the estimated values (eta), pure residual (res), standardized residual (resstd), standardized weighted residual 2 (resstd2), residual deviance (resdeviance), Cooks distance (cookD) and generalized leverage (glbp).
- `log_likelihood` - Log-likelihood of the fitted model.
- `aicc` - Corrected Akaike information criterion.
- `r2` - Pseudo R2 and adjusted pseudo R2 statistics.
- `bp_test` - Breusch-Pagan test for heteroscedasticity.
- `link_function` - The link function used in modeling.
- `n_iter` - Number of iterations used in convergence.

Examples

```
data(saopaulo)
output_list=betareg_gwbr("prop_landline",c("prop_urb","prop_poor"),saopaulo)

## Parameters
output_list$parameter_estimates

## R2 and AICc
output_list$r2
output_list$aicc
```

---

gss_gwbr   *Golden Section Search Algorithm*

Description

The Golden Section Search (GSS) algorithm is used in searching for the best bandwidth for geographically weighted regression. For more details see Da Silva and Mendes (2018).

Usage

```
gss_gwbr(
  yvar,
  xvar,
  lat,
  long,
  data,
  method = c("fixed_g", "fixed_bsq", "adaptive_bsq"),
  link = c("logit", "probit", "loglog", "cloglog"),
  type = c("cv", "aic"),
  globalmin = TRUE,
  distancekm = TRUE,
  maxint = 100
)
```

Arguments

yvar	A vector with the response variable name.
xvar	A vector with descriptive variable(s) name(s).
lat	A vector with the latitude variable name.
long	A vector with the longitude variable name.
data	A data set object with yvar and xvar.
method	Kernel function used to set bandwidth parameter. The options are: "fixed_g", "fixed_bsq" or "adaptive_bsq". The default is "fixed_g".
link	The link function used in modeling. The options are: "logit", "probit", "loglog" or "cloglog". The default is "logit".

type	Can be ″cv″, when the Cross-Validation function is used to estimate the bandwidth or ″aic″, when the AIC function is used. The default is ″cv″.
globalmin	Logical. If TRUE search for the global minimum. The default is TRUE.
distancekm	Logical. If TRUE use the distance in kilometers otherwise, use the Euclidean distance. The default is TRUE.
maxint	A maximum number of iterations to numerically maximize the log-likelihood function in search of parameter estimates. The default is maxint=100.

Value

A list that contains:

- global_min - Global minimum of the function, giving the best bandwidth (h).
- local_mins - Local minimums of the function.
- type - Function used to estimate the bandwidth.

Examples

```
data(saopaulo)
output_list=gss_gwbr("prop_landline",c("prop_urb","prop_poor"),"y","x",saopaulo,"fixed_g")

## Best bandwidth
output_list$global_min
```

---

gwbr   *Geographically Weighted Beta Regression*

Description

Fits a local regression model for each location using the beta distribution, recommended for rates and proportions, using a parametrization with mean (transformed by the link function) and precision parameter (called phi). For more details see Da Silva and Lima (2017).

Usage

```
gwbr(
  yvar,
  xvar,
  lat,
  long,
  h,
  data,
  xglobal = NA_character_,
  grid = data.frame(),
  method = c("fixed_g", "fixed_bsq", "adaptative_bsq"),
```

```
    link = c("logit", "probit", "loglog", "cloglog"),
    distancekm = TRUE,
    global = FALSE,
    maxint = 100
)
```

Arguments

yvar	A vector with the response variable name.
xvar	A vector with descriptive variable(s) name(s).
lat	A vector with the latitude variable name.
long	A vector with the longitude variable name.
h	The bandwidth parameter.
data	A data set object with yvar and xvar.
xglobal	A vector with descriptive variable(s) name(s) with global effect.
grid	A data set with the location variables. Only used when the location variable are in another data set, different from data set used in parameter data. Variable name "lat" is expected for latitude and "long" for longitude.
method	The kernel function used. The options are: "fixed_g", "fixed_bsq" or "adaptive_bsq". The default is "fixed_g".
link	The link function used in modeling. The options are: "logit", "probit", "loglog" or "cloglog". The default is "logit".
distancekm	Logical. If TRUE use the distance in kilometers otherwise, use the Euclidean distance. The default is TRUE.
global	Logical. If TRUE return to global model, giving the results from betareg_gwbr function. The default is FALSE.
maxint	A maximum number of iterations to numerically maximize the log-likelihood function in search of the parameter estimates. The default is maxint=100.

Value

A list that contains:

- parameter_estimates_qtls - Parameter estimates quartiles and interquartile range.
- parameter_estimates_desc - Parameter estimates mean, minimum and maximum.
- std_qtls - Standard deviation quartiles and interquartile range.
- std_desc - Standard deviation mean, minimum and maximum.
- est_n_parameters - Number of parameters.
- est_gwr_parameters - Effective number of parameters in the local model.
- phi - Vector of precision parameter estimates.
- global_parameter - Global parameter estimates, when existing.
- global_phi - Global scale parameter estimate, when existing.
- global_parameter_tab - Global parameter estimates table, when existing.

- `residuals` - Table with observed values (`y`), estimated values (`yhat`), the link function applied in the estimated values (`eta`), pure residual (`res`), standardized residual (`resstd`), standardized weighted residual 2 (`resstd2`), residual deviance (`resdeviance`), Cooks distance (`cookD`), generalized leverage (`glbp`) and number of iterations (`iteration`).

- `log_likelihood` - Log-likelihood of the fitted model.

- `aicc` - Corrected Akaike information criterion.

- `r2` - Pseudo R2 and adjusted pseudo R2 statistics.

- `bp_test` - Breusch-Pagan test for heteroscedasticity.

- `w` - Matrix of weights.

- `parameters` - Table with parameter estimates of each model.

- `significance` - Significance level of each model.

- `bandwidth` - Bandwidth used.

- `link_function` - The link function used in modeling.

Examples

```
data(saopaulo)
output_list=gwbr("prop_landline",c("prop_urb", "prop_poor"),"y","x",116.3647,saopaulo)

## Descriptive statistics of the parameter estimates
output_list$parameter_estimates_desc

## Table with all parameter estimates and your respective statistics
output_list$parameters
```

---

saopaulo   *Sao Paulo dataset*

Description

Data from 2010 of the municipalities of Sao Paulo state, Brazil.

Usage

```
data(saopaulo)
```

Format

A data frame with 644 observations and 14 variables:

municipality Municipality name.

state State.

geocode Municipality geocode according to IBGE.

households Number of households.

landline Number of households with landline.

pop Total population.

pop_rural Rural population.

pop_urb Urban population.

hdim Municipal Human Development Index.

prop_urb Proportion of urban population.

prop_poor Proportion of poor population (Considering per capita household income equal or less than R$140.00 per month).

prop_landline Proportion of households with landline.

x Longitude of the centroid of the city.

y Latitude of the centroid of the city.

Index