Package ‘geessbin’

September 9, 2024

Type Package

Title Modified Generalized Estimating Equations for Binary Outcome

Version 1.0.0

Maintainer Ryota Ishii <r.ishii0808@gmail.com>

Description Analyze small-sample clustered or longitudinal data with binary outcome using modified generalized estimating equations (GEE) with bias-adjusted covariance estimator. The package provides any combination of three GEE methods and 12 covariance estimators.

Depends R (>= 3.5.0)

Imports MASS (>= 7.3-45)

License GPL (>= 2)

Encoding UTF-8

LazyData true

URL https://github.com/rtishii/geessbin

RoxygenNote 7.3.2

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Ryota Ishii [aut, cre],
Tomohiro Ohigashi [ctb],
Kazushi Maruo [ctb],
Masahiko Gosho [ctb]

Repository CRAN

Date/Publication 2024-09-08 23:10:02 UTC

Contents

---

geessbin *Modified Generalized Estimating Equations for Binary Outcome*

Description

geessbin analyzes small-sample clustered or longitudinal data using modified generalized estimating equations (GEE) with bias-adjusted covariance estimator. This function assumes binary outcome and uses the logit link function. This function provides any combination of three GEE methods (conventional and two modified GEE methods) and 12 covariance estimators (unadjusted and 11 bias-adjusted estimators).

Usage

```
geessbin(
  formula,
  data = parent.frame(),
  id = NULL,
  corstr = "independence",
  repeated = NULL,
  beta.method = "PGEE",
  SE.method = "MB",
  b = NULL,
  maxitr = 50,
  tol = 1e-05,
  scale.fix = FALSE,
  conf.level = 0.95
)
```

Arguments

formula	Object of class formula: symbolic description of model to be fitted (see documentation of lm and formula for details).
data	Data frame.
id	Vector that identifies the subjects or clusters (NULL by default).
corstr	Working correlation structure. The following are permitted: "independence", "exchangeable", "ar1", and "unstructured" ("independence" by default).
repeated	Vector that identifies repeatedly measured variable within each subject or cluster. If repeated = NULL, as is the case in function gee, data are assumed to be sorted so that observations on a cluster are contiguous rows for all entities in the formula.
beta.method	Method for estimating regression parameters (see Details section). The following are permitted: "GEE", "PGEE", and "BCGEE" ("PGEE" by default).
SE.method	Method for estimating standard errors (see Details section). The following are permitted: "SA", "MK", "KC", "MD", "FG", "PA", "GS", "MB", "WL", "WB", "FW", and "FZ" ("MB" by default).

b	Numeric vector specifying initial values of regression coefficients. If b = NULL (default value), the initial values are calculated using the ordinary or Firth logistic regression assuming that all the observations are independent.
maxitr	Maximum number of iterations (50 by default).
tol	Tolerance used in fitting algorithm (1e-5 by default).
scale.fix	Logical variable; if TRUE, the scale parameter is fixed at 1 (FALSE by default).
conf.level	Numeric value of confidence level for confidence intervals (0.95 by default).

Details

Details of beta.method are as follows:

- "GEE" is the conventional GEE method (Liang and Zeger, 1986)
- "BCGEE" is the bias-corrected GEE method (Paul and Zhang, 2014; Lunardon and Scharfstein, 2017)
- "PGEE" is the bias reduction of the GEE method obtained by adding a Firth-type penalty term to the estimating equation (Mondol and Rahman, 2019)

Details of SE.method are as follows:

- "SA" is the unadjusted sandwich variance estimator (Liang and Zeger, 1986)
- "MK" is the MacKinnon and White estimator (MacKinnon and White, 1985)
- "KC" is the Kauermann and Carroll estimator (Kauermann and Carroll, 2001)
- "MD" is the Mancl and DeRouen estimator (Mancl and DeRouen, 2001)
- "FG" is the Fay and Graubard estimator (Fay and Graubard, 2001)
- "PA" is the Pan estimator (Pan, 2001)
- "GS" is the Gosho et al. estimator (Gosho et al., 2014)
- "MB" is the Morel et al. estimator (Morel et al., 2003)
- "WL" is the Wang and Long estimator (Wang and Long, 2011)
- "WB" is the Westgate and Burchett estimator (Westgate and Burchett, 2016)
- "FW" is the Ford and Wastgate estimator (Ford and Wastgate, 2018)
- "FZ" is the Fan et al. estimator (Fan et al., 2013)

Descriptions and performances of some of the above methods can be found in Gosho et al. (2023).

Value

The object of class "geessbin" representing the results of modified generalized estimating equations with bias-adjusted covariance estimators. Generic function summary provides details of the results.

References

- Fan, C., Zhang, D., and Zhang, C. H. (2013). A comparison of bias-corrected covariance estimators for generalized estimating equations. *Journal of Biopharmaceutical Statistics*, 23, 1172–1187, doi:10.1080/10543406.2013.813521.

- Fay, M. P. and Graubard, B. I. (2001). Small-sample adjustments for Wald-type tests using sandwich estimators. *Biometrics*, 57, 1198–1206, doi:10.1111/j.0006341X.2001.01198.x.

- Ford, W. P. and Westgate, P. M. (2018). A comparison of bias-corrected empirical covariance estimators with generalized estimating equations in small-sample longitudinal study settings. *Statistics in Medicine*, 37, 4318–4329, doi:10.1002/bimj.201600182.

- Gosho, M., Ishii, R., Noma, H., and Maruo, K. (2023). A comparison of bias-adjusted generalized estimating equations for sparse binary data in small-sample longitudinal studies. *Statistics in Medicine*, 42, 2711–2727, doi:10.1002/sim.9744.

- Gosho, M., Sato, T., and Takeuchi, H. (2014). Robust covariance estimator for small-sample adjustment in the generalized estimating equations: A simulation study. *Science Journal of Applied Mathematics and Statistics*, 2, 20–25, doi:10.11648/j.sjams.20140201.13.

- Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96, 1387–1396, doi:10.1198/016214501753382309.

- Liang, K. and Zeger, S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13–22, doi:10.1093/biomet/73.1.13.

- Lunardon, N. and Scharfstein, D. (2017). Comment on ‘Small sample GEE estimation of regression parameters for longitudinal data’. *Statistics in Medicine*, 36, 3596–3600, doi:10.1002/sim.7366.

- MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29, 305–325, doi:10.1016/03044076(85)901587.

- Mancl, L. A. and DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*, 57, 126–134, doi:10.1111/j.0006341X.2001.00126.x.

- Mondol, M. H. and Rahman, M. S. (2019). Bias-reduced and separation-proof GEE with small or sparse longitudinal binary data. *Statistics in Medicine*, 38, 2544–2560, doi:10.1002/sim.8126.

- Morel, J. G., Bokossa, M. C., and Neerchal, N. K. (2003). Small sample correlation for the variance of GEE estimators. *Biometrical Journal*, 45, 395–409, doi:10.1002/bimj.200390021.

- Pan, W. (2001). On the robust variance estimator in generalised estimating equations. *Biometrika*, 88, 901–906, doi:10.1093/biomet/88.3.901.
- Paul, S. and Zhang, X. (2014). Small sample GEE estimation of regression parameters for longitudinal data. *Statistics in Medicine*, 33, 3869–3881, doi:10.1002/sim.6198.
- Wang, M. and Long, Q. (2011). Modified robust variance estimator for generalized estimating equations with improved small-sample performance. *Statistics in Medicine*, 30, 1278–1291, doi:10.1002/sim.4150.
- Westgate, P. M. and Burchett, W. W. (2016). Improving power in small-sample longitudinal studies when using generalized estimating equations. *Statistics in Medicine*, 35, 3733–3744, doi:10.1002/sim.6967.

Examples

```
data(wheeze)

# analysis of PGEE method with Morel et al. covariance estimator
res <- geessbin(formula = Wheeze ~ City + factor(Age), data = wheeze, id = ID,
                corstr = "ar1", repeated = Age, beta.method = "PGEE",
                SE.method = "MB")

# hypothesis tests for regression coefficients
summary(res)
```

`geessbin_all` *Function for analysis using all combinations of GEE methods and covariance estimators*

Description

`geessbin_all` provides analysis results using all combinations of three GEE methods and 12 covariance estimators.

Usage

```
geessbin_all(
  formula,
  data = parent.frame(),
  id = NULL,
  corstr = "independence",
  repeated = NULL,
  b = NULL,
  maxitr = 50,
  tol = 1e-05,
```

```
  scale.fix = FALSE,
  conf.level = 0.95
)
```

Arguments

`formula`	Object of class formula: symbolic description of model to be fitted (see documentation of `lm` and `formula` for details).
`data`	Data frame.
`id`	Vector that identifies the subjects or clusters (`NULL` by default).
`corstr`	Working correlation structure. The following are permitted: `"independence"`, `"exchangeable"`, `"ar1"`, and `"unstructured"` (`"independence"` by default).
`repeated`	Vector that identifies repeatedly measured variable within each subject or cluster. If `repeated = NULL`, as is the case in function gee, data are assumed to be sorted so that observations on a cluster are contiguous rows for all entities in the formula.
`b`	Numeric vector specifying initial values of regression coefficients. If `b = NULL` (default value), the initial values are calculated using the ordinary or Firth logistic regression assuming that all the observations are independent.
`maxitr`	Maximum number of iterations (50 by default).
`tol`	Tolerance used in fitting algorithm (`1e-5` by default).
`scale.fix`	Logical variable; if `TRUE`, the scale parameter is fixed at 1 (`FALSE` by default).
`conf.level`	Numeric value of confidence level for confidence intervals (0.95 by default).

Value

The list containing two data frames. The first is a table of estimates of regression coefficients, standard errors, z-values, and p-values. The second is a table of odds ratios and confidence intervals.

---

sqrtmat *Square root of nonsymmetric matrix*

Description

`sqrtmat` is used to calculate the square root of $E_i - H_{ii}$, which is an adjustment factor in Kauermann and Carroll-type method.

Usage

```
sqrtmat(M)
```

Arguments

`M`	Square matrix whose square root is to be computed.

Value

The square root of `M`

References

Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96, 1387–1396, doi:10.1198/016214501753382309.

`wheeze` *Wheeze dataset*

Description

The data studied the effect of air pollution on the health of 16 children. The outcome variable was the wheezing status measured consistently four times yearly at ages of 9, 10, 11, and 12 years.

Format

A data frame with 64 observations on the following 6 variables:

`ID` child identifier.

`Wheeze` binary indicator of wheezing presence.

`City` binary indicator of whether the child lives in Kingston (0 = Portage; 1 = Kingston).

`Age` age of child in years ranging from 9 to 12.

`Smoke` measure of smoking habits (cigarettes per day) of child's mother.

References

- Hardin, J. and Hilbe, J. (2013). *Generalized Estimating Equations, 2nd edition*. Chapman and Hall, London.
- Lipsitz, S. R., Fitzmaurice, G. M., Orav, E. J., and Laird, N. M. (1994). Performance of Generalized Estimating Equations in Practical Situations. *Biometrics*, 50, 270–278, doi:10.2307/2533218.

Examples

```
data(wheeze)
```

Index