Package 'CIMPLE'

July 21, 2025

Title Analysis of Longitudinal Electronic Health Record (EHR) Data with Possibly Informative Observational Time

Version 0.1.0

Description Analyzes longitudinal Electronic Health Record (EHR) data with possibly informative observational time. These methods are grouped into two classes depending on the inferential task. One group focuses on estimating the effect of an exposure on a longitudinal biomarker while the other group assesses the impact of a longitudinal biomarker on time-to-diagnosis outcomes. The accompanying paper is Du et al (2024) <doi:10.48550/arXiv.2410.13113>.

License GPL (>= 3)

Depends R (>= 2.10)

Imports dplyr, JMbayes2, lme4, mice, nleqslv, nlme, statmod, stats, survival, utils

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Author Jiacong Du [aut] (ORCID: <https://orcid.org/0000-0002-7768-9083>), Howard Baik [cre] (ORCID: <https://orcid.org/0009-0000-8942-1618>)

Maintainer Howard Baik <howard.baik@yale.edu>

Repository CRAN

Date/Publication 2024-11-12 13:20:10 UTC

Contents

long_data *long_data*

Description

A subset of data from the World Health Organization Global Tuberculosis Report ...

Usage

```
long_data
```

Format

who:
A data frame with 7,240 rows and 60 columns:

country Country name

iso2, iso3 2 & 3 letter ISO country codes

year Year ...

Source

https://www.who.int/teams/global-tuberculosis-programme/data

long_est *Coefficient estimation in the longitudinal model*

Description

This function offers a collection of methods of coefficient estimation in a longitudinal model with possibly informative observation time. These methods include Standard linear mixed-effect model (standard_LME), Linear mixed-effect model adjusted for the historical number of visits (VA_LME), Joint model of the visiting process and the longitudinal process accounting for measured confounders (JMVL_LY), Inverse-intensity-rate-ratio weighting approach (IIRR_weighting), Joint model of the visiting process and the longitudinal process with dependent latent variables (JMVL_Liang), Imputation-based approach with linear mixed-effect model (imputation_LME), and Joint model of the visiting process and the longitudinal process with a shared random intercept (JMVL_G).

Usage

```
long_est(
  long_data,
  method,
  id_var,
  outcome_var,
  LM_fixedEffect_variables = NULL,
```

```
  time = NULL,
  LM_randomEffect_variables = NULL,
  VPM_variables = NULL,
  imp_time_factor = NULL,
  optCtrl = list(method = "nlminb", kkt = FALSE, tol = 0.2, maxit = 20000),
  control = list(verbose = FALSE, tol = 0.001, GHk = 10, maxiter = 150),
  ...
)
```

Arguments

`long_data`	Longitudinal dataset
`method`	The following methods are available: • `standard_LME`: Standard linear mixed-effect model. • `VA_LME`: Linear mixed-effect model adjusted for the historical number of visits. • `JMVL_LY`: Joint model of the visiting process and the longitudinal process accounting for measured confounders. • `IIRR_weighting`: Inverse-intensity-rate-ratio weighting approach. • `JMVL_Liang`: Joint model of the visiting process and the longitudinal process with dependent latent variables. • `imputation_LME`: Imputation-based approach with linear mixed-effect model. • `JMVL_G`: Joint model of the visiting process and the longitudinal process with a shared random intercept.
`id_var`	Variable for the subject ID to indicate the grouping structure.
`outcome_var`	Variable name for the longitudinal outcome variable.
`LM_fixedEffect_variables`	Vector input of variable names with fixed effects in the longitudinal model. Variables should not contain `time`.
`time`	Variable for the observational time.
`LM_randomEffect_variables`	Vector input of variable names with random effects in the longitudinal model. This argument is `NULL` for methods including `JMVL_LY`, `JMVL_G` and `IIRR_weighting`.
`VPM_variables`	Vector input of variable names in the visiting process model.
`imp_time_factor`	Scale factor for the time variable. This argument is only needed in the imputation-based methods i.e., `imputation_LME`.
`optCtrl`	Control parameters for running the mixed-effect model. See the `control` argument in `lme4::lmer()`.
`control`	Control parameters for the `JMVL_G` method: • `verbose`: `TRUE` or `FALSE` for outputting checkpoint after each iteration. Default is `FALSE`. • `tol`: Tolerance for convergence. • `GHk`: Number of gaussian-hermite quadrature points. Default is `10`. • `maxiter`: Maximum number of iteration. Default is `150`.
`...`	Additional arguments to `nleqslv::nleqslv()`.

Value

`beta_hat`: Estimated coefficients in the longitudinal model.

Other output in each method:

- `standard_LME`:
  - `beta_sd`: Standard deviation of the estimated coefficients.
- `VA_LME`:
  - `beta_sd`: Standard deviation of the estimated coefficients.
- `JMVL_LY`:
  - `gamma_hat`: Estimated coefficients in the visiting process model.
- `IIRR_weighting`:
  - `gamma_hat`: Estimated coefficients in the visiting process model.
- `JMVL_Liang`:
  - `gamma_hat`: Estimated coefficients in the visiting process model.

References

Buzkova, P. and Lumley, T. (2007). Longitudinal data analysis for generalized linear models with follow-up dependent on outcome-related variables. Canadian Journal of Statistics, 35(4):485–500.

Gasparini, A., Abrams, K. R., Barrett, J. K., Major, R. W., Sweeting, M. J., Brunskill, N. J., and Crowther, M. J. (2020). Mixed-effects models for health care longitudinal data with an informative visiting process: A monte carlo simulation study. Statistica Neerlandica, 74(1):5–23.

Liang, Y., Lu, W., and Ying, Z. (2009). Joint modeling and analysis of longitudinal data with informative observation times. Biometrics, 65(2):377–384.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. Journal of the American Statistical Association, 96(453):103–126.

Examples

```
# Setup arguments
train_data

time_var = "time"
id_var = "id"
outcome_var = "Y"
VPM_variables = c("Z", "X")
LM_fixedEffect_variables = c("Z", "X")
LM_randomEffect_variables = "Z"

# Run the standard LME model
fit_standardLME = long_est(long_data=train_data,
                           method="standard_LME",
                           id_var=id_var,
                           outcome_var=outcome_var,
                           LM_fixedEffect_variables = LM_fixedEffect_variables,
                           time = time_var,
```

```
                          LM_randomEffect_variables = LM_randomEffect_variables,
                          VPM_variables = VPM_variables)
# Return the coefficient estimates
fit_standardLME$beta_hat

# Run the VA_LME model
fit_VALME = long_est(long_data=train_data,
                     method="VA_LME",
                     id_var=id_var,
                     outcome_var=outcome_var,
                     LM_fixedEffect_variables = LM_fixedEffect_variables,
                     time = time_var,
                     LM_randomEffect_variables = LM_randomEffect_variables,
                     VPM_variables = VPM_variables)
# Return the coefficient estimates
fit_VALME$beta_hat
```

---

surv_data *long_data*

Description

A subset of data from the World Health Organization Global Tuberculosis Report ...

Usage

```
surv_data
```

Format

who**:**

A data frame with 7,240 rows and 60 columns:

country Country name

iso2, iso3 2 & 3 letter ISO country codes

year Year ...

Source

https://www.who.int/teams/global-tuberculosis-programme/data

surv_est *Coefficient estimation in the survival model with longitudinal measurements.*

Description

This function offers a collection of methods of coefficient estimation in a survival model with a longitudinally measured predictor. These methods include Cox proportional hazard model with time-varying covariates (cox), Joint modeling the longitudinal and disease diagnosis processes (JMLD), Joint modeling the longitudinal and disease diagnosis processes with an adjustment for the historical number of visits in the longitudinal model (VA_JMLD), Cox proportional hazard model with time-varying covariates after imputation (Imputation_Cox), Cox proportional hazard model with time-varying covariates after imputation with an adjustment for the historical number of visits in the longitudinal model (VAImputation_Cox).

Usage

```
surv_est(
  long_data,
  surv_data,
  method,
  id_var,
  time = NULL,
  survTime = NULL,
  survEvent = NULL,
  LM_fixedEffect_variables = NULL,
  LM_randomEffect_variables = NULL,
  SM_timeVarying_variables = NULL,
  SM_timeInvariant_variables = NULL,
  imp_time_factor = NULL
)
```

Arguments

long_data	Longitudinal dataset.
surv_data	Survival dataset.
method	The following methods are available:

- cox: Cox proportional hazard model with time-varying covariates.
- JMLD: Joint modeling the longitudinal and disease diagnosis processes.
- VA_JMLD: Joint modeling the longitudinal and disease diagnosis processes with an adjustment for the historical number of visits in the longitudinal model.
- Imputation_Cox: Cox proportional hazard model with time-varying covariates after imputation.

- VAImputation_Cox: Cox proportional hazard model with time-varying covariates after imputation with an adjustment for the historical number of visits in the longitudinal model.

id_var Variable for the subject ID to indicate the grouping structure.

time Variable for the observational time.

survTime Variable for the survival time.

survEvent Variable for the survival event.

LM_fixedEffect_variables
Vector input of variable names with fixed effects in the longitudinal model. Variables should not contain time.

LM_randomEffect_variables
Vector input of variable names with random effects in the longitudinal model.

SM_timeVarying_variables
Vector input of variable names for time-varying variables in the survival model.

SM_timeInvariant_variables
Vector input of variable names for time-invariant variables in the survival model.

imp_time_factor
Scale factor for the time variable. This argument is only needed in the imputation-based methods, e.g., Imputation_Cox and VAImputation_Cox. The default is NULL (no scale).

Value

alpha_hat: Estimated coefficients for the survival model.

Other output in each method:

- JMLD:
  - beta_hat: Estimated coefficients for the longitudinal model.
- VA_JMLD:
  - beta_hat: Estimated coefficients for the longitudinal model.

References

Rizopoulos, D. (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. Journal of statistical software, 35:1–33.

Rizopoulos, D. (2012b). Joint models for longitudinal and time-to-event data: With applications in R. CRC press.

Examples

```
# Setup arguments

id_var = "id"
time = "time"
survTime = "D"
survEvent = "d"
LM_fixedEffect_variables = c("Age","Sex","SNP")
```

```
LM_randomEffect_variables = c("SNP")
SM_timeVarying_variables = c("Y")
SM_timeInvariant_variables = c("Age","Sex","SNP")
imp_time_factor = 1

# Run the cox model
fit_cox = surv_est(surv_data = surv_data,
                   long_data = long_data,
                   method = "cox",
                   id_var = id_var,
                   time = time,
                   survTime = survTime,
                   survEvent = survEvent,
                   LM_fixedEffect_variables = LM_fixedEffect_variables,
                   LM_randomEffect_variables = LM_randomEffect_variables,
                   SM_timeVarying_variables = SM_timeVarying_variables,
                   SM_timeInvariant_variables = SM_timeInvariant_variables,
                   imp_time_factor = imp_time_factor)
# Return the coefficient estimates
fit_cox$alpha_hat
```

`train_data` *long_data*

Description

A subset of data from the World Health Organization Global Tuberculosis Report ...

Usage

```
train_data
```

Format

who**:**

A data frame with 7,240 rows and 60 columns:

country Country name

iso2, iso3 2 & 3 letter ISO country codes

year Year ...

Source

https://www.who.int/teams/global-tuberculosis-programme/data

Index