Package: marlod (via r-universe)

August 28, 2024

Type Package

Title Marginal Modeling for Exposure Data with Values Below the LOD

Version 0.1.1

Depends R (>= 2.10)

Imports MASS, miWQS, quantreg, stats, knitr

Maintainer I-Chen Chen <okv0@cdc.gov>

Description Functions of marginal mean and quantile regression models are used to analyze environmental exposure and biomonitoring data with repeated measurements and non-detects, as well as longitudinal exposure data that include non-detects and time-dependent covariates.

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.1

NeedsCompilation yes

Repository CRAN

LazyData true

Suggests rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Author I-Chen Chen [cre, aut] (<https://orcid.org/0000-0001-6764-8395>), Philip Westgate [ctb], Liya Fu [ctb]

Date/Publication 2024-08-27 22:00:02 UTC

Contents

Fillin   *Fill-in or Substitution Methods*

Description

Uses substitution methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD).

Usage

```
Fillin(y, lod, substitue)
```

Arguments

y	A list of numeric values or a vector of the observed values.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", and "QQplot".

Details

Single value imputation techniques, such as LOD/2 or LOD/$\sqrt{2}$ ("LOD2" or "LODS2") (Hornung and Reed, 1990; Burstyn and Teschke, 1999), and β-substitution method ("BetaMean" and "BetaGM") (Ganser and Hewett, 2010), are used to assign a value to a range between 0 and the LOD. "QQplot" represents the multiple order value imputation technique that depicts the natural logarithm of the uncensored or detected observed values versus the Z-scores and fits a linear regression presented in a quantile-quantile (QQ) plot (Pleil, 2016).

Value

A list of numeric values or a vector with imputed values that are assigned to non-detects.

Author(s)

I-Chen Chen

References

Burstyn, I., Teschke, K. (1999). Studying the determinants of exposure: a review of methods. *American Industrial Hygiene Association Journal*, 60, 57–72.

Ganser, G. H., Hewett, P. (2010). An accurate substitution method for analyzing censored data. *Journal of Occupational and Environmental Hygiene*, 7, 233–44.

Hornung, R. W., Reed, L. D. (1990). Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*, 5, 46–51.

Pleil, J. D. (2016). QQ-plots for assessing distributions of biomarker measurements and generating defensible summary statistics. *Journal of Breath Research*, 10, 035001.

Examples

```
## Uses an example from Ganser and Hewett (2010).
library(marlod)

y <- c(0,0,0,3.06,4.41,7.23,8.29,9.52,19.94,20.25) #LOD=3
lod <- 3

Fillin(y, lod, "None")

Fillin(y, lod, "LOD")

Fillin(y, lod, "LOD2")

Fillin(y, lod, "LODS2")

Fillin(y, lod, "BetaMean")

Fillin(y, lod, "BetaGM")

Fillin(y, lod, "QQplot")
```

MGEE   *Function of a Generalized Estimating Equation (GEE) Model*

Description

The function is used to calculate a empirical MSE minimization criterion (EMMC) value in the "Selected.GEE" function.

Usage

```
MGEE(id, y, x, lod, substitue, corstr, typetd, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
typetd	An atomic vector specifying the types of time-dependent covaraites. The length of this vector is the number of regression paramenters, including the intercept. "1" is assigned to any time-indepednet covariates or covariates in a cluster study.
maxiter	The maximum number of iterations.

Value

An object of class "MGEE".

Modified.GEE	*Function of a Generalized Estimating Equation (GEE) Model*

Description

Runs a marginal mean regression model using generalized estimating equation (GEE) estimation method for repeated measures data with values less than the limit of detection (LOD).

Usage

```
Modified.GEE(id, y, x, lod, substitue, corstr, typetd, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".

corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
typetd	An atomic vector specifies the types of time-dependent covaraites, with the length of the vector equal to the number of regression parameters, excluding the intercept. For time-independent covariates or those in a cluster study, "1" is assigned.
maxiter	The maximum number of iterations.

Details

The function modifies the supplementary R function for GEE in Westgate (2014a), in whcih small-sample standard error corrections are applied (Kauermann and Carroll, 2001; Mancl and DeRouen, 2001; Westgate, 2013). More discussions about the use of covariance corrections can be found in Westgate (2016), and Ford and Westgate (2017, 2018). With the marginal modeling, Chen *et al.* (2024) incorporate the fill-in methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD) are assigned values. This function also presents the results of the "trace of the empirical covariance matrix" (TECM) (Westgate, 2014b) and the "correlation information criterion" (CIC) (Hin and Wang, 2009). Both criteria have been shown to be preferable to other criteria in choosing an analysis method and corresponding structure (Westgate, 2014a).

See the Details of the "Fillin" function for introduction of the available fill-in or substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM") (Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata15", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata15".

Value

An object of class "Modified.GEE" representing the fit.

Note

The function is capable of analyzing one measurement or more than one repeated measurements per subject. Unbalanced repeated measurements are also permittable.

Author(s)

Philip M. Westgate and I-Chen Chen

References

Chen, I-C., Bertke, S. J., Estill, C. F. (2024). Compare the Marginal Effects for Environmental Exposure and Biomonitoring Data with Repeated Measurements and Values Below the Limit of Detection. *Journal of Exposure Science and Environmental Epidemiology*. doi:10.1038/s41370-024006407

Ford, W. P., Westgate, P. M. (2017). Improved standard error estimator for maintaining the validity of inference in cluster randomized trials with a small number of clusters. *Biometrical Journal*, 59, 478–95.

Ford, W. P., 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, 4318–29.

Hin, L. Y., Wang, Y.-G. (2009). Working-correlation-structure identification in generalized estimating equations. *Statistics in Medicine*, 28, 642–658.

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

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

Mancl, L. A., DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*, 57, 126–134.

Westgate, P. M. (2013). A bias correction for covariance estimators to improve inference with generalized estimating equations that use an unstructured correlation matrix. *Statistics in Medicine*, 32, 2850–2858.

Westgate, P. M. (2014a). Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. *Biometrical Journal*, 56, 461–476.

Westgate, P. M. (2014b). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine*, 33, 2222–2237.

Westgate, P. M. (2016). A covariance correction that accounts for correlation estimation to improve finite-sample inference with generalized estimating equations: a study on its applicability with structured correlation matrices. *Journal of Statistical Computation and Simulation*, 86, 1891–1900.

See Also

Selected.GEE.

Examples

```
## Uses the simdata15 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)

data(simdata15)

id=as.matrix(as.vector(t(simdata15$id)))
y=as.matrix(as.vector(t(simdata15$y)))
x1=as.matrix(as.vector(t(simdata15$x1)))
x2=as.matrix(as.vector(t(simdata15$x2)))
x=cbind(x1,x2)

## LOD=2 is equivalent to detection proportion=56.3% (censoring proportion=43.7%).
lod=2
```

```
## Intercept is not included in the "x"
Modified.GEE(id, y, x, lod, "None", "exchangeable", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "LOD", "AR-1", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "LOD2", "exchangeable", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "LODS2", "AR-1", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "BetaMean", "exchangeable", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "BetaGM", "AR-1", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "MIWithID", "exchangeable", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "MIWithIDRM", "AR-1", c(1,1), 1000)

Modified.GEE(id, y, x, lod, "QQplot", "exchangeable", c(1,1), 1000)
```

Modified.GMM   *Function of a Generalized Method of Moments Model*

Description

Runs a marginal mean regression model using generalized method of moments (GMM) estimation method for repeated measures data with values less than the limit of detection (LOD).

Usage

```
Modified.GMM(id, y, x, lod, substitue, beta, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
beta	A matrix of initial parameter estimates, e.g., these estimates could be from general linear model or generalized estimating equation (GEE) using independence working structure.
maxiter	The maximum number of iterations.

Details

The modified GMM approach was originally proposed by Chen and Westgate (2017), in whcih a linear shrinkage method of Han and Song (2011) was incorporated to resolve potential singularity problems. The method should be utilized when the Moore–Penrose generalized inverse fails to solve the weighting matrix. Small-sample standard error corrections were also applied to the modified GMM (Mancl and DeRouen, 2001; Westgate, 2012). With the marginal modeling, Chen *et al.* (2024) incorporate the fill-in methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD) are assigned values. This function also presents the results of the "trace of the empirical covariance matrix" (TECM) (Westgate, 2014a) and the "correlation information criterion" (CIC) (Hin and Wang, 2009). Both criteria have been shown to be preferable to other criteria in choosing an analysis method and corresponding structure (Westgate, 2014b).

See the Details of the "Fillin" function for introduction of the available fill-in or substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM") (Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata15", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata15".

Value

An object of class "Modified.GMM" representing the fit.

Author(s)

I-Chen Chen

References

Chen, I-C., Bertke, S. J., Estill, C. F. (2024). Compare the Marginal Effects for Environmental Exposure and Biomonitoring Data with Repeated Measurements and Values Below the Limit of Detection. *Journal of Exposure Science and Environmental Epidemiology*. doi:10.1038/s41370-024006407

Chen, I-C., Westgate, P. M. (2017). Improved methods for the marginal analysis of longitudinal data in the presence of timedependent covariates. *Statistics in Medicine*, 36, 2533–2546.

Han, P., Song, P. X. K. (2011). A note on improving quadratic inference functions using a linear shrinkage approach. *Statistics and Probability Letters*, 81, 438–445.

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

Mancl, L. A., DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*, 57, 126–134.

Westgate, P. M. (2012). A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine*, 31, 4003–4022.

Westgate, P. M. (2014a). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine*, 33, 2222–2237.

Westgate, P. M. (2014b). Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. *Biometrical Journal*, 56, 461–476.

Examples

```
## Uses the simdata15 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)

data(simdata15)

id=as.matrix(as.vector(t(simdata15$id)))
y=as.matrix(as.vector(t(simdata15$y)))
x1=as.matrix(as.vector(t(simdata15$x1)))
x2=as.matrix(as.vector(t(simdata15$x2)))
x=cbind(x1,x2)

## LOD=2 is equivalent to detection proportion=56.3% (censoring proportion=43.7%).
lod=2

## Gets initial estimates for the GMM approach through independence structure
initial=glm(y ~ x1 + x2, data=simdata15, family=gaussian)
beta_initial=as.matrix(initial$coefficients)

## Intercept is not included in the "x"
Modified.GMM(id, y, x, lod, "None", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "LOD", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "LOD2", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "LODS2", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "BetaMean", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "BetaGM", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "MIWithID", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "MIWithIDRM", beta_initial, 1000)

Modified.GMM(id, y, x, lod, "QQplot", beta_initial, 1000)
```

Modified.QIF   *Function of a Quadratic Inference Function (QIF) Model*

Description

Runs a marginal mean regression model using quadratic inference function (QIF) estimation method for repeated measures data with values less than the limit of detection (LOD).

Usage

```
Modified.QIF(id, y, x, lod, substitue, corstr, beta, typetd, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
beta	A matrix of initial parameter estimates, e.g., these estimates could be from general linear model or generalized estimating equation (GEE) using independence working structure.
typetd	An atomic vector specifies the types of time-dependent covaraites, with the length of the vector equal to the number of regression parameters, excluding the intercept. For time-independent covariates or those in a cluster study, "1" is assigned.
maxiter	The maximum number of iterations.

Details

The function modifies the supplementary R function for GEE in Westgate (2014a), in whcih small-sample standard error corrections are applied (Kauermann and Carroll, 2001; Mancl and DeRouen, 2001; Westgate and Braun, 2012; Westgate, 2012, 2014b). With the marginal modeling, Chen *et al.* (2024) incorporate the fill-in methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD) are assigned values. This function also presents the results of the "trace of the empirical covariance matrix" (TECM) (Westgate, 2014c) and the "correlation information criterion" (CIC) (Hin and Wang, 2009). Both criteria have been shown to be preferable to other criteria in choosing an analysis method and corresponding structure (Westgate, 2014a).

See the Details of the "Fillin" function for introduction of the available fill-in or substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM")

(Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata15", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata15".

Value

An object of class "Modified.QIF" representing the fit.

Note

The function is capable of analyzing one measurement or more than one repeated measurements per subject. Unbalanced repeated measurements are also permittable.

Author(s)

Philip M. Westgate and I-Chen Chen

References

Chen, I-C., Bertke, S. J., Estill, C. F. (2024). Compare the Marginal Effects for Environmental Exposure and Biomonitoring Data with Repeated Measurements and Values Below the Limit of Detection. *Journal of Exposure Science and Environmental Epidemiology*. doi:10.1038/s41370-024006407

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

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

Mancl, L. A., DeRouen, T. A. (2001). A covariance estimator for GEE with improved small-sample properties. *Biometrics*, 57, 126–134.

Westgate, P. M., Braun, T. M. (2012). The effect of cluster size imbalance and covariates on the estimation performance of quadratic inference functions. *Statistics in Medicine*, 31, 2209–2222.

Westgate, P. M. (2012). A bias-corrected covariance estimate for improved inference with quadratic inference functions. *Statistics in Medicine*, 31, 4003–4022.

Westgate, P. M. (2014a). Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. *Biometrical Journal*, 56, 461–476.

Westgate, P. M. (2014b). A comparison of utilized and theoretical covariance weighting matrices on the estimation performance of quadratic inference functions. *Communications in Statistics – Simulation and Computation*, 43, 2432–2443.

Westgate, P. M. (2014c). Improving the correlation structure selection approach for generalized estimating equations and balanced longitudinal data. *Statistics in Medicine*, 33, 2222–2237.

See Also

Selected.QIF.

Examples

```
## Uses the simdata15 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)

data(simdata15)

id=as.matrix(as.vector(t(simdata15$id)))
y=as.matrix(as.vector(t(simdata15$y)))
x1=as.matrix(as.vector(t(simdata15$x1)))
x2=as.matrix(as.vector(t(simdata15$x2)))
x=cbind(x1,x2)

## LOD=2 is equivalent to detection proportion=56.3% (censoring proportion=43.7%).
lod=2

## Gets initial estimates for the QIF approach through independence structure
initial=glm(y ~ x1 + x2, data=simdata15, family=gaussian)
beta_initial=as.matrix(initial$coefficients)

## Intercept is not included in the "x"
Modified.QIF(id, y, x, lod, "None", "exchangeable", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "LOD", "AR-1", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "LOD2", "exchangeable", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "LODS2", "AR-1", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "BetaMean", "exchangeable", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "BetaGM", "AR-1", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "MIWithID", "exchangeable", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "MIWithIDRM", "AR-1", beta_initial, c(1,1), 1000)

Modified.QIF(id, y, x, lod, "QQplot", "exchangeable", beta_initial, c(1,1), 1000)
```

MQIF   *Function of a Quadratic Inference Function (QIF) Model*

Description

The function is used to calculate a empirical MSE minimization criterion (EMMC) value in the "Selected.QIF" function.

Usage

```
MQIF(id, y, x, lod, substitue, corstr, beta, typetd, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
beta	A matrix of initial parameter estimates, e.g., these estimates could be from general linear model or generalized estimating equation (GEE) using independence working structure.
typetd	An atomic vector specifying the types of time-dependent covaraites. The length of this vector is the number of regression paramenters, including the intercept. "1" is assigned to any time-indepednet covariates or covariates in a cluster study.
maxiter	The maximum number of iterations.

Value

An object of class "MQIF".

Quantile.FWZ *Function of a Quantile Regression Model*

Description

Runs a marginal quantile regression model for repeated measures data with values less than the limit of detection (LOD).

Usage

```
Quantile.FWZ(y, x, lod, substitue, tau, corstr, typetd, data)
```

Arguments

y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).

substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
tau	A numeric value of quantile level, e.g., tau=0.25 for 25th quantile and tau=0.5 for median.
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
typetd	An atomic vector specifies the types of time-dependent covaraites, with the length of the vector equal to the number of regression parameters, excluding the intercept. For time-independent covariates or those in a cluster study, "1" is assigned.
data	A data frame that originazes the given data into two-dimensional structure of rows and columns.

Details

This function modifies the R functions provided by Dr. Liya Fu and based on the manuscript of Fu *et al.* (2015). Chen *et al.* (2021) further applied the Gaussian pseudolikelihood approach for quantile regression to environmental exposure and biomonitoring repeated measures data with values less than the limit of detection (LOD). Fill-in or substitution methods, including single and multiple value imputation techniques, were used to assign values for non-detects.

See the Details of the "Fillin" function for introduction of the available substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM") (Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata15", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata15".

Value

An object of class "Quantile.FWZ" representing the fit.

Author(s)

Liya Fu and I-Chen Chen

References

Chen, I-C., Bertke, S. J., Curwin, B. D. (2021). Quantile regression for exposure data with repeated measures in the presence of non-detects. *Journal of Exposure Science and Environmental Epidemiology*, 31, 1057–1066.

Fu, L., Wang, Y.-G., Zhu, M. (2015). A Gaussian pseudolikelihood approach for quantile regression with repeated measurements. *Computational Statistics and Data Analysis*, 84, 41–53.

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

See Also

Quantile.select.FWZ.

Examples

```
## Uses the simdata15 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)
library(quantreg)

data(simdata15)

y=as.matrix(as.vector(t(simdata15$y)))
x1=as.matrix(as.vector(t(simdata15$x1)))
x2=as.matrix(as.vector(t(simdata15$x2)))
x=cbind(matrix(1,length(x1),1),x1,x2)

## LOD=2 is equivalent to detection proportion=50% (censoring proportion=50%).
lod=2

## Median or 50th quantile is given.
tau=0.5

## Examples to perform the function
Quantile.FWZ(y, x, lod, "BetaGM", tau, "AR-1", c(1,1), simdata15)

Quantile.FWZ(y, x, lod, "QQplot", tau, "exchangeable", c(1,1), simdata15)

Quantile.FWZ(y, x, lod, "MIWithID", tau, "exchangeable", c(1,1), simdata15)
```

Quantile.select.FWZ   *Function to Select a Type of Time-Dependent Covaraite Through a Quantile Regression Model*

Description

Selects a type of time-dependent covaraite through a marginal quantile regression model for longitudinal exposure data with values less than the limit of detection (LOD).

Usage

```
Quantile.select.FWZ(y, x, lod, substitue, tau, data)
```

Arguments

y   A column matrix of the observed outcome values or responses.

x   A matrix of covariate values, for which the number of columns is the number of covariates.

lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
tau	A numeric value of quantile level, e.g., tau=0.25 for 25th quantile and tau=0.5 for median.
data	A data frame that originazes the given data into two-dimensional structure of rows and columns.

Details

The function modifies the R functions provided by Dr. Liya Fu and based on the manuscript of Fu *et al.* (2015). Chen *et al.* (2024) further applied the Gaussian pseudolikelihood approach for quantile regression to environmental exposure and biomonitoring longitudinal data with values less than the limit of detection (LOD) and time-dependent covaraites. The work to select a working type of time-dependent covaraite is based on the manuscript of Chen and Westgate (2021).

Fill-in or substitution methods, including single and multiple value imputation techniques, were used to assign values for non-detects. See the Details of the "Fillin" function for introduction of the available substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM") (Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata58", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata58".

Value

An object of class "Quantile.select.FWZ" representing the fit.

Author(s)

Liya Fu and I-Chen Chen

References

Chen, I-C., Bertke, S. J., Dahm, M. M. (2024). Quantile regression for longitudinal data with values below the limit of detection and time-dependent covariates – application to modeling carbon nanotube and nanofiber exposures. *Annals of Work Exposures and Health*. doi:10.1093/annweh/wxae068

Chen, I-C., Westgate, P. M. (2021). Marginal quantile regression for longitudinal data analysis in the presence of time-dependent covariates. *The International Journal of Biostatistics*, 17, 267–282.

Fu, L., Wang, Y.-G., Zhu, M. (2015). A Gaussian pseudolikelihood approach for quantile regression with repeated measurements. *Computational Statistics and Data Analysis*, 84, 41–53.

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

See Also

[Quantile.FWZ](#).

Examples

```
## Uses the simdata58 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)
library(quantreg)

data(simdata58)

y=as.matrix(as.vector(t(simdata58$y)))
x1=as.matrix(as.vector(t(simdata58$x1)))
x=cbind(matrix(1,length(x1),1),x1)

## LOD=0.5 is equivalent to detection proportion=50.7% (censoring proportion=49.3%).
lod=0.5

## Median or 50th quantile is given.
tau=0.5

## Examples to perform the function
Quantile.select.FWZ(y, x, lod, "BetaMean", tau, simdata58)

Quantile.select.FWZ(y, x, lod, "QQplot", tau, simdata58)

Quantile.select.FWZ(y, x, lod, "MIWithID", tau, simdata58)
```

Selected.GEE	*Function to Select a Type of Time-Dependent Covaraite Through a Generalized Estimating Equation Model*

Description

Selects a type of time-dependent covaraite through a marginal mean regression model using generalized estimating equation (GEE) estimation method for longitudinal exposure data with values less than the limit of detection (LOD).

Usage

```
Selected.GEE(id, y, x, lod, substitue, corstr, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.

x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
maxiter	The maximum number of iterations.

Details

The function modifies the supplementary R function for GEE in Westgate (2014). With the marginal modeling, Chen *et al.* (2024) incorporate the fill-in methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD) are assigned values. Based on the manuscripts of Chen and Westgate (2017, 2019), this function also enable to use a empirical MSE minimization criterion (EMMC) to select a working type of time-dependent covaraite.

See the Details of the "Fillin" function for introduction of the available fill-in or substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM") (Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the covariate, e.g., "id in "simdata58", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata58".

Value

An object of class "Selected.GEE" representing the fit.

Note

The function is capable of analyzing one measurement or more than one repeated measurements per subject. Unbalanced repeated measurements are also permittable.

Author(s)

Philip M. Westgate and I-Chen Chen

References

Chen, I.-C., Bertke, S. J., Estill, C. F. (2024). Compare the Marginal Effects for Environmental Exposure and Biomonitoring Data with Repeated Measurements and Values Below the Limit of Detection. *Journal of Exposure Science and Environmental Epidemiology*. doi:10.1038/s41370-024006407

Chen, I.-C., Westgate, P. M. (2017). Improved methods for the marginal analysis of longitudinal data in the presence of time-dependent covariates. *Statistics in Medicine*, 36, 2533–46.

Chen, I-C., Westgate, P. M. (2019). A novel approach to selecting classification types for time-dependent covariates in the marginal analysis of longitudinal data. *Statistical Methods in Medical Research*, 28, 3176–86.

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

Westgate, P. M. (2014). Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. *Biometrical Journal*, 56, 461–476.

See Also

Modified.GEE, MGEE.

Examples

```
## Uses the simdata58 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)

data(simdata58)

id=as.matrix(as.vector(t(simdata58$id)))
y=as.matrix(as.vector(t(simdata58$y)))
x1=as.matrix(as.vector(t(simdata58$x1)))

## LOD=0.5 is equivalent to detection proportion=50.7% (censoring proportion=49.3%).
lod=0.5

## Intercept is not included in the "x1"
Selected.GEE(id, y, x1, lod, "None", "exchangeable", 1000)

Selected.GEE(id, y, x1, lod, "LOD", "AR-1", 1000)

Selected.GEE(id, y, x1, lod, "LOD2", "exchangeable", 1000)

Selected.GEE(id, y, x1, lod, "LODS2", "AR-1", 1000)

Selected.GEE(id, y, x1, lod, "BetaMean", "exchangeable", 1000)

Selected.GEE(id, y, x1, lod, "BetaGM", "AR-1", 1000)

Selected.GEE(id, y, x1, lod, "MIWithID", "exchangeable", 1000)

Selected.GEE(id, y, x1, lod, "MIWithIDRM", "AR-1", 1000)

Selected.GEE(id, y, x1, lod, "QQplot", "exchangeable", 1000)
```

Selected.QIF	*Function to Select a Type of Time-Dependent Covaraite Through a Quadratic Inference Function Model*

Description

Selects a type of time-dependent covaraite through a marginal quantile regression model using quadratic inference function (QIF) estimation method for longitudinal exposure data with values less than the limit of detection (LOD).

Usage

```
Selected.QIF(id, y, x, lod, substitue, corstr, beta, maxiter)
```

Arguments

id	A column matrix of subject IDs. The number of rows is the total number of observations. Data must be sorted by IDs.
y	A column matrix of the observed outcome values or responses.
x	A matrix of covariate values, for which the number of columns is the number of covariates.
lod	A numeric value of limit of detection (LOD).
substitue	A character string specifying the substitution approach, including "None", "LOD", "LOD2", "LODS2", "BetaMean", "BetaGM", "MIWithID", "MIWithIDRM", and "QQplot".
corstr	A character string specifying the working correlation structure, given by either "exchangeable" or "AR-1".
beta	A matrix of initial parameter estimates, e.g., these estimates could be from general linear model or generalized estimating equation (GEE) using independence working structure.
maxiter	The maximum number of iterations.

Details

The function modifies the supplementary R function for QIF in Westgate (2014). With the marginal modeling, Chen *et al.* (2024) incorporate the fill-in methods, including single and multiple value imputation techniques, such that any measurements less than the limit of detection (LOD) are assigned values. Based on the manuscripts of Chen and Westgate (2017, 2019), this function also enable to use a empirical MSE minimization criterion (EMMC) to select a working type of time-dependent covaraite.

See the Details of the "Fillin" function for introduction of the available fill-in or substitution methods. For a multiple random value imputation technique, it provides an alternative for environmental exposure and biomonitoring data with non-detects, in which the imputed values can be generated using a regression of an exposure measurement on covariate(s) ("MIWithID" and "MIWithIDRM")

(Lubin *et al.*, 2004). Information of identification (ID) would be included in "MIWithID" as the co-variate, e.g., "id in "simdata58", while ID and order of cluster size or time points would be treated as the covariates in "MIWithIDRM", e.g. "id" and "visit" in "simdata58".

Value

An object of class "Selected.QIF" representing the fit.

Note

The function is capable of analyzing one measurement or more than one repeated measurements per subject. Unbalanced repeated measurements are also permittable.

Author(s)

Philip M. Westgate and I-Chen Chen

References

Chen, I-C., Bertke, S. J., Estill, C. F. (2024). Compare the Marginal Effects for Environmental Exposure and Biomonitoring Data with Repeated Measurements and Values Below the Limit of Detection. *Journal of Exposure Science and Environmental Epidemiology*. doi:10.1038/s41370-024006407

Chen, I-C., Westgate, P. M. (2017). Improved methods for the marginal analysis of longitudinal data in the presence of time-dependent covariates. *Statistics in Medicine*, 36, 2533–46.

Chen, I-C., Westgate, P. M. (2019). A novel approach to selecting classification types for time-dependent covariates in the marginal analysis of longitudinal data. *Statistical Methods in Medical Research*, 28, 3176–86.

Lubin, J. H., Colt, J. S., Camann, D., *et al.* (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environmental Health Perspectives*, 112, 1691–6.

Westgate, P. M. (2014). Criterion for the simultaneous selection of a working correlation structure and either generalized estimating equations or the quadratic inference function approach. *Biometrical Journal*, 56, 461–476.

See Also

Modified.QIF, MQIF.

Examples

```
## Uses the simdata58 to run the marginal models.
library(marlod)
library(MASS)
library(miWQS)

data(simdata58)

id=as.matrix(as.vector(t(simdata58$id)))
y=as.matrix(as.vector(t(simdata58$y)))
```

```
x1=as.matrix(as.vector(t(simdata58$x1)))

## LOD=0.5 is equivalent to detection proportion=50.7% (censoring proportion=49.3%).
lod=0.5

## Gets initial estimates for the QIF approach through independence structure
initial=glm(y ~ x1, data=simdata58, family=gaussian)
beta_initial=as.matrix(initial$coefficients)

## Intercept is not included in the "x1"
Selected.QIF(id, y, x1, lod, "None", "exchangeable", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "LOD", "AR-1", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "LOD2", "exchangeable", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "LODS2", "AR-1", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "BetaMean", "exchangeable", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "BetaGM", "AR-1", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "MIWithID", "exchangeable", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "MIWithIDRM", "AR-1", beta_initial, 1000)

Selected.QIF(id, y, x1, lod, "QQplot", "exchangeable", beta_initial, 1000)
```

---

simdata15   *Simulated Dataset 15*

Description

The 15th dataset from the simulation study has 100 subjects (sample size is 30). Each subject has three repeated measurements. The independent variables or covariates are simulated from a binomial distribution with a parameter value of $p = 0.5$ and a uniform distribution $U(0, 1)$, respectively. Correlated errors for models with repeated measures are accounted for and assumed to follow a multivariate normal distribution, $MVN(0, R(\alpha))$. A first-order autoregressive (AR-1) correlation structure with a correlation parameter of $\alpha = 0.7$ is incorporated into the multivariate normal distribution. The true values of 1, 1, and 1 are corresponded to the marginal intercept and two slopes, accordingly.

Usage

```
data("simdata15")
```

Format

A data frame with 30 subjects and each subject has three repeated measurements, i.e., number of cluster size or time points. A list that contains two variables:

y A column matrix of the continuous outcome values.

int A column matrix of the intercept values of one.

x1 A column matrix of the binary covariate values that follow a binomial distribution.

x2 A column matrix of the continuous covariate values that follow a uniform distribution.

id A column matrix of the numbers of identification.

visit A column matrix of the order of cluster size or time points.

Examples

```
library(marlod)
data(simdata15)
```

```
simdata58                    Simulated Dataset 58
```

Description

The 58th dataset from the simulation study has 100 subjects (sample size is 100). Each subject has three repeated measurements. Detailed model mechanism can be found in the setting II for type III time-dependent covariate on page 90 of Lai and Small (2007). The two random effects in the mechanism are mutually independent and normally distributed with mean 0 and variances 1. The true values of 0 and 0.69 are corresponded to the marginal intercept and slope, accordingly.

Usage

```
data("simdata58")
```

Format

A data frame with 100 subjects and each subject has three repeated measurements, i.e., number of cluster size or time points. A list that contains one variable:

y A column matrix of the continuous outcome values.

int A column matrix of the intercept values of one.

x1 A column matrix of the binary covariate values that follow a binomial distribution.

id A column matrix of the numbers of identification.

visit A column matrix of the order of cluster size or time points.

References

Lai, T.L., Small, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. *Journal of the Royal Statistical Society: Series B*, 69, 79–99.

Examples

```
library(marlod)
data(simdata58)
```

Index