**mimix- reference based multiple imputation for sensitivity analysis of a longitudinal clinical trial with protocol deviation.**

# **Description**

See github.com/UCL/mimix for project description, source code and help files.

mimix imputes missing numerical outcomes for a longitudinal trial with protocol deviation under distinct treatment arm-based assumptions for the unobserved data, following the general algorithm of Carpenter, Roger, and Kenward (2013).

# **Installation**

*At the moment* download individual files functions.R and main.R direct from github, saving into suitable project directories.

Go to the program on github then click Alt-Blame simultaneously

# **Usage**

Preliminaries

1. Ensure R environment empty
2. Edit the source command in main.R to point to the directory containing the functions.R file
3. Import data, typically from a csv file,
4. recode treatment variable if necessary to be numeric, starting from 1 .

Corresponding R commands (example for asthma data)

1. rm(list = ls())
2. source("N:/Documents/GitHub/mimix/mimixR/functions.R")
3. read.table("http://raw.githubusercontent.com/UCL/mimix/master/data/asthma.csv",header=TRUE,sep=",",fileEncoding = "UTF-8-BOM")
4. mxdata$treat<-ifelse(mxdata$treat==2,1,ifelse(mxdata$treat==3,2,9))

# **Running the imputation**

#Runmimix generic arguments

Runmimix(covar,depvar,treatvar,idvar,timevar,M=1,refer,meth,seedval=101,priorvar, burnin=1000,bbetween=NULL,methodindiv)

#Runmimix specific to Asthma data

**mimix\_outputlist<-**Runmimix(c("base"),"fev","treat","id","time",100,1,"J2R",101,"jeffreys",1000,NULL,NULL)

#get the M imputed datasets

# impdatasets<-getimpdatasets(mimix\_outputlist**)**

#run regression on imputed datasets combining using Rubin’s rules

regressimp(impdatasets,"fev12~treat+base")

# **Arguments**

Runmimix(covar,depvar,treatvar,idvar,timevar,M=1,refer,meth,seedval=101,priorvar, burnin=1000,bbetween=NULL,methodindiv)

Covar A vector of type character specifying baseline covariates (including the dependent baseline

depvar A character specifying the numeric outcome variable with missing data

treatvar A character specifying the treatment group variable, whose value must be recoded to numeric if not already numeric.

idvar A character specifying individuals

timevar A character specifying numeric units of time (or visits) between observations, assumed equidistant

M A number specifying number of imputations

refer character specifying the level of treatment group to act as the reference group mandatory with j2r, cir and cr methods of imputation

meth A character specifying imputation method from 5 (see below)

seedval A number to set the random number seed

prior A character to set the prior used in the norm2 mcmc

burnin A number specify the number of iterations for the burn-in period in the mcmc procedure

bbetween A number specifying number of iterations between imputations in the MCMC procedure

methodindiv A vector of type character specifying the variable containing the individual-specific imputation method together with the level of treatment chosen for the reference group for each individual ,mandatory for j2r, cir, cr.

# **Details**

The program is a port from the Stata mimix program. It reads data in long format with one individual per timepoint . The data is transformed into wide data with the depvar transformed into the same number of variables as there are time-points.

### Algorithm for obtaining imputed data sets

mimix implements the general algorithm of Carpenter, Roger, and Kenward (2013)

1. Transform data into wide and find missing data patterns

2. For each treatment arm, take all the observed data, assume MAR, and fit an MVN distribution with an unstructured covariance

3. Draw a mean vector and covariance matrix from the posterior distribution for each treatment arm, we use the MCMC method to draw from the appropriate Bayesian posterior.

4. Use the draws in step 2 to form the joint distribution for each deviating individual's observed and missing outcome data as required, depending on the method chosen for the missing data.

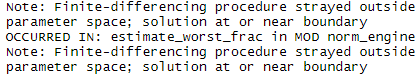
5. Construct the conditional distribution of missing given observed outcome data for each individual who deviated from the joint distribution formed in step 3. Sample missing data from the conditional distributions to create a completed dataset.

6. Repeat steps 3-5 m times, resulting in m imputed datasets.

8. Run models on the m imputed data sets combining results using Rubin’s rules.

The MCMC method chosen uses functions emNorm and mcmcNorm from the norm2 package by Schafer (norm2 user guide 2016).

Warning msgs from norm2 are explained as difficulties with noninformative priors – the emNorm algorithm did converge but the procedure for estimating the worst fraction and worst linear function failed.



The warnng indicates the Emo solution to be near the boundary of the parameter space implying the estimate of the covariance matrix to be nearly singular. The warning messages may disappear if the ridge prior is used.

Covariates and base value depvar assumed completely observed .

**Examples**

**# accupuncture data**

mxdata<-read.table("http://raw.githubusercontent.com/UCL/mimix/master/data/accupuncture.csv",header=TRUE,sep=",",fileEncoding = "UTF-8-BOM")

#recode treatment groups to 1,2..

mxdata$treat<-ifelse(mxdata$treat=="control",1,ifelse(mxdata$treat=="accupuncture",2,9))

mimix\_outputlist <- Runmimix(c("head\_base","sex","age"),"head","treat","id","time",100,1,"MAR",201,c("jeffreys"),1000,NULL,NULL)

impdatasets<-getimpdatasets(mimix\_outputlist)

regressimp(impdatasets,"head12~treat+head\_base+age+sex")

**#anti-depressant data with methodvar & referencevar**

mxdata<-read.table("http://raw.githubusercontent.com/UCL/mimix/master/data/SASantidepmethodvar.csv",header=TRUE,sep=",",fileEncoding = "UTF-8-BOM")

mxdata$TREATMENT.NAME<-ifelse(mxdata$TREATMENT.NAME=="PLACEBO",1,ifelse(mxdata$TREATMENT.NAME=="DRUG",2,9))

mimix\_outputlist <- Runmimix(c("basval"),"change","TREATMENT.NAME","PATIENT.NUMBER","VISIT.NUMBER",100,NULL,NULL,101,c("jeffreys"),1000,NULL,c("methodvar","referencevar"))

impdatasets<-getimpdatasets(mimix\_outputlist)

regressimp(impdatasets,"change4~TREATMENT.NAME+basval")

# **Testing output**

MAR

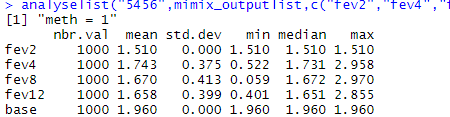
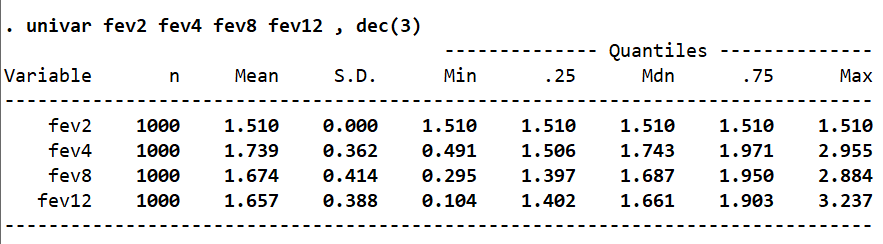


Figure 1 mimix fev treat, id(id) time(time) method(mar) covariates(base) clear m(1000) seed(101)



CR

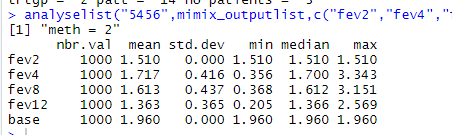
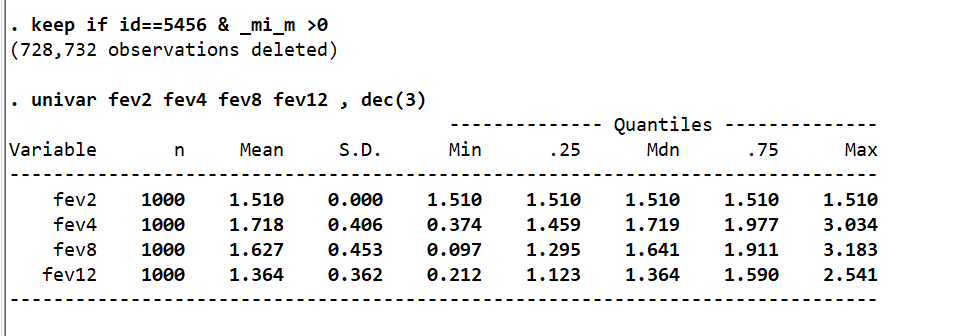


Figure 2 mimix fev treat, id(id) time(time) method(CR) refgroup(2) covariates(base) clear m(100



CIR

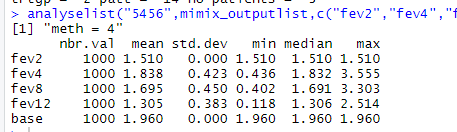
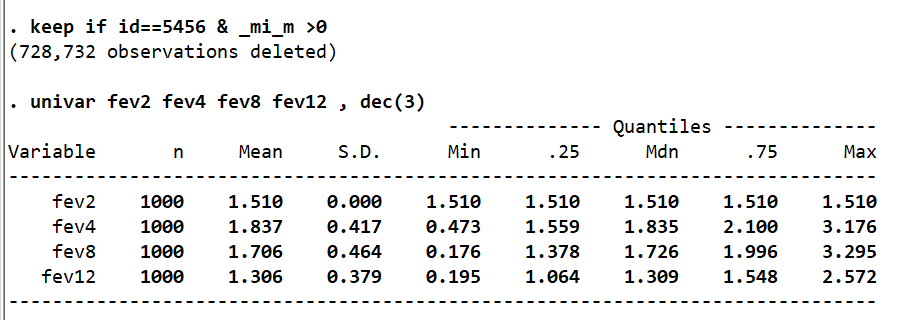
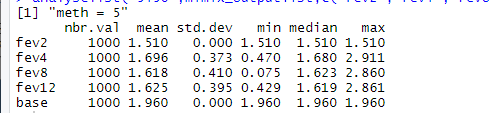
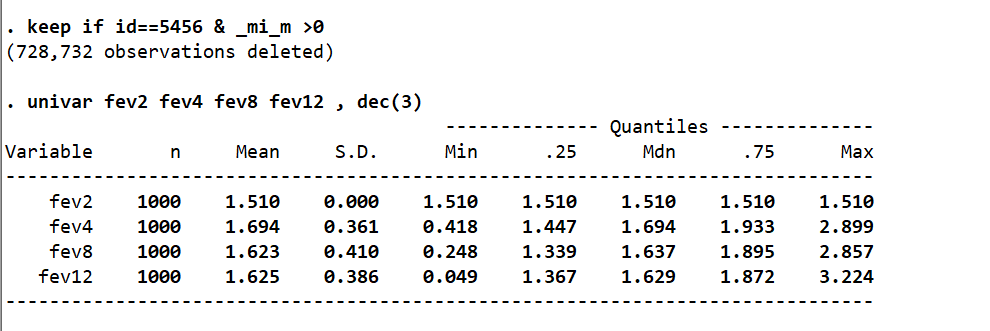


Figure 3 mimix fev treat, id(id) time(time) method(CIR) refgroup(2) covariates(base) clear m(1000) seed(101)



LMCF





Anti-depressant data, imp method for each individual

Figure 4 Runmimix(c("basval"),"change","TREATMENT.NAME","PATIENT.NUMBER","VISIT.NUMBER",1000,NULL,NULL,101,c("jeffreys"),1000,NULL,c("methodvar","referencevar"))

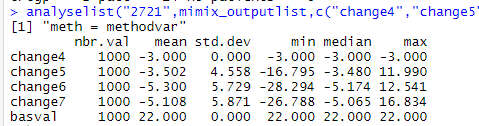


Figure 5 mimix change tmentno,id(patientnumber) time(visitnumber) covar(basval ) methodvar(methodvar) refgroupvar

