The missed values in data may be an obstacle to the using of certain statistics methods for analysis that are not able to process them. A large number of missed values ignored during statistical analysis, or incorrect methods of filling them can lead to erroneous conclusions and thereby negate the whole study.

The module aimv (analysis and imputation of missing values) is intended for solve the problems of analyzing missed values and their imputation in the language of programming Python. It contains 20 functions:

aggr – calculate or plot the amount of missing values in each variable and the amount of missing values in certain combinations of variables;

bar\_miss – barplot with highlighting of missing values in other variable for the studied variable;

chols – return submatrix of theta corresponding to the columns of mc;

em\_norm – performs maximum-likelihood estimation on the matrix of incomplete data using the EM algorithm;

gauss – generates random numbers from a Gaussian distribution;

getparam\_norm – takes a parameter vector, such as one produced by em\_norm and returns a list of parameters on the original scale;

gtmc – finds the column numbers of the missing variables, and stores them in the first nmc elements of mc. Does not go beyond last column;

gtoc – finds the column numbers of the observed variables, and stores them in the first noc elements of oc. Does not go beyond last column;

hist\_miss – histogram with highlighting of missing values in variables by splitting each bin into two parts;

imp\_norm – draws missing elements of a data matrix under the multivariate normal model and a user-supplied parameter;

initn – initializes theta;

is1n – performs I-step of data augmentation. Randomly draws missing data Xmis given theta, and stores the sufficient statistics in t. Theta must be in sweep(0) condition. Answer is returned in unswept condition;

mcar\_test – use Little’s (1988) test statistic to assess if data is missing completely at random (MCAR);

object\_miss – for obtain the number of the rows in a data frame that have a "large" number of missing values. "Large" can be defined either as a proportion of the number of columns or as the number in itself;

prelim\_norm – perform preliminary manipulations on matrix of continuous data. Rows are sorted by missing data pattern;

rangen – generates random numbers for a Gaussian distribution;

sigex – extracts submatrix of theta corresponding to the columns of mc;

swp – performs sweep on a symmetric matrix in packed storage. Sweeps on pivot position. Sweeps only the (0:submat,0:submat) submatrix. If dir=1, performs ordinary sweep. If dir=-1, performs reverse sweep;

swpobs – sweeps theta to condition on the observed variables;

tobsn – tabulates the known part of the sscp matrix for all missingness patterns.