

Penerapan metode sequence regression multivariate imputation pada data multivariat normal = Application of sequence regression multivariate imputation method on multivariate normal data

Nurzaman, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20496317&lokasi=lokal>

Abstrak

Pada setiap analisis statistik memungkinkan berhadapan dengan missing values atau missing data karena pada saat survei kemungkinan ada responden yang tidak dapat menjawab pertanyaan atau tidak ingin menjawab pertanyaan pada saat wawancara survei. Missing values tidak dapat langsung dilakukan analisis menggunakan analisis data lengkap, oleh karena itu missing values telah menjadi masalah yang sering dihadapi oleh para peneliti. Dataset survei biasanya terdiri dari sejumlah besar variabel kontinu salah satunya berdistribusi multivariat normal. Salah satu cara untuk menangani missing values dapat dilakukan dengan imputasi, yaitu proses pengisian atau penggantian missing values pada dataset dengan nilai-nilai yang mungkin berdasarkan informasi yang didapatkan pada dataset tersebut. Penelitian ini akan menerapkan metode sequence regression multivariate imputation (SRMI) untuk imputasi missing values pada data multivariat normal.

SRMI merupakan metode imputasi ganda yang nilai imputasinya didapatkan dari model sequence of regression yaitu setiap variabel yang mengandung missing values diregresikan terhadap semua variabel lain yang tidak mengandung missing values sebagai variabel prediktor. Cara mendapatkan nilai imputasi digunakan pendekatan iterasi untuk menarik nilai dari distribusi posterior prediktif pada missing values di bawah masing-masing model regresi secara beruntun. Penelitian ini menggunakan data multivariat normal yang telah dibangkitkan sebanyak 500 observasi dengan menggunakan lima nilai imputasi ganda dan hasil evaluasi kualitas imputasi menggunakan Root Mean Square Error (RMSE). Hasil evaluasi kualitas imputasi dapat dikatakan baik jika nilai RMSE semakin kecil, maka eror semakin kecil atau nilai estimasi mendekati nilai sebenarnya (Chai & Draxler, 2014) dan hasil yang didapatkan nilai RMSE kecil sehingga SRMI dapat diterapkan untuk melakukan imputasi terhadap data multivariat normal.

.....Missing values are the absence of data items for an observation or more observations that can result in the loss of certain information. During surveys, there are often missing values or missing data because there are likely respondents who cannot answer the question or do not want to answer the question. That is a problem for researchers because, with missing values, the results of observation cannot be analyzed properly. Survey datasets usually consist of continuous variables, one of which is a normal multivariate distribution. One way to deal with missing values can be done by imputation, which is the process of filling or replacing missing values in a dataset with possible values based on the information obtained in the dataset. This study will apply the sequence regression multivariate imputation (SRMI) method for missing values imputation in normal multivariate data.

SRMI is a multiple imputation method whose implication value is obtained from the sequence of regression model, that is, every variable containing missing values is regressed on all other variables that do not contain missing values as predictor variables. The method of obtaining imputation values is used by the iterative approach to drawing values from the predictive posterior distribution in the missing values below each successive regression model. This study uses multivariate normal data that has been generated a total of 500

observations using five multiple imputation values and the evaluation results using Root Mean Square Error (RMSE) which have little value in applying to normal multivariate data so SRMI can be applied to impute normal multivariate data.