

ABSTRAK

Judul Tesis	:	Imputasi <i>Missing Value</i> pada Data yang Mengandung <i>Outlier</i> .
Kata Kunci	:	<i>Missing Data, Outlier, Predictive Mean Matching, Least Trimmed Squares, Robust Estimation</i>
Nama	:	Hafti Mardiah
NPM	:	140720090012
Program Studi	:	Statistika Terapan
Bidang Kajian Utama	:	Statistika Sosial
Tim Pembimbing	:	1. Gandhi Pawitan, Ph.D. 2. Budhi Handoko, M.Si.
Tahun Kelulusan	:	2010

Abstrak

Missing Data merupakan salah satu permasalahan yang sering terjadi pada sebuah survei. Imputasi adalah pilihan penanganan *missing data* yang paling bijak dari pada membuang sebagian observasi atau variabel yang mengandung *missing value*, mengingat bahwa data sangat mahal dan berharga. Penanganan *missing value* pada sekumpulan data yang terdapat *outlier* menjadi perhatian khusus karena sebagian besar metode imputasi dengan mekanisme *Missing at Random* (MAR) dan *Missing Completely at Random* (MCAR) mengasumsikan data berdistribusi normal multivariat. Asumsi ini menjadi tidak valid ketika terdapat *outlier* pada data, sehingga sebaiknya menggunakan metode imputasi berdasarkan estimasi yang *robust* terhadap *outlier*. Metode *Predictive Mean Matching* (PMM) adalah salah satu alternatif metode imputasi komposit, penggabungan metode imputasi regresi

dengan metode imputasi *nearest neighbour*, yang mengasumsikan data berasal dari distribusi normal multivariat. Ketika asumsi normalitas dilanggar, PMM menghasilkan nilai hasil imputasi yang tidak masuk akal dan statistik *Efficiency Relative* yang lebih rendah dibandingkan dengan metode imputasi regresi *Least Trimmed Squares* (LTS). Metode imputasi regresi LTS merupakan penggabungan algoritma LTS dan algoritma imputasi regresi.

Abstract

Missing data is the most frequent problem that occurs in a survey. Thus, imputation is a prudent alternative of handling the missing data instead of reducing the number of observations or variables due to its cost achieved and value. The treatment of the missing data in the presence of outlier becomes the major problem which is the most imputation method based on the Missing at Random (MAR) and Missing Completely at Random (MCAR) mechanism. Moreover, it assumes data originated from a multivariate normal distribution, which is no longer valid in the presence of outliers in the data. For instance, Predictive Mean Matching (PMM), a combination of regression imputation method and the nearest neighbour method, assumes the data originated from a multivariate normal distribution. When the normality assumption is violated, the predictive mean matching method does not yield plausible imputed values plus the performance of the Efficiency Relative (ER) is below compared to the ER of Least Trimmed Squares (LTS) regression imputation method. LTS regression imputation method is actually a regression imputation method which its parameter is the result of LTS regression estimation then combined with the regression imputation algorithm.