

ABSTRACT**COMPARISON OF REGRESSION IMPUTATION AND MULTIVARIATE IMPUTATION BY CHAINED EQUATION TO ESTABLISH A MODEL****(Factors Influencing the Timing of First Sexual Intercourse Among Adolescent Based on Survey of RPJMN 2015)**

Data collection activities have a higher risk of missing data. Missing data may produce biased estimates and standard errors were greater, so the imputation technique is needed. The aim of this study was analyze and compare between regression imputation and multivariate imputation by chained equation (MICE) method then establish a model of factors influencing the timing of first sexual intercourse among adolescent based on Survey of RPJMN 2015..

This study was non-reactive study and used raw data taken from BKKBN East Java Province. The population was all respondents of Survei RPJMN 2015 which had dating and pre-marital sexual intercourse. Simulation data of this study was from complete data which eliminated as much as 5%, 10%, and 15%. The comparative parameters was MSE, GFI, and RMSEA.

Friedman Test analysis showed that there was no different between the imputed data and the original data. Based on MSE analysis, MICE better than Regression Imputation. Based on GFI and RMSEA values, Regression Imputation as good as MICE in missing data 5% and 10%, but in missing data 10% MICE better than Regression Imputation.

The conclusion was MICE better than Regression Imputation. Hopefully, this result could help other researchers, especially for BKKBN, not to deleting unit which has missing data. The advantages of the MICE method are applicable to not only for normality data, and can be applied to vacancies up to 15%.

Keywords: Missing Data, Regression Imputation, MICE, Path Analysis

ABSTRAK

**PERBANDINGAN HASIL *REGRESSION IMPUTATION* DAN
MULTIVARIATE IMPUTATION BY CHAINED EQUATION
PADA PEMBENTUKAN MODEL STRUKTURAL**

**(Analisis Faktor yang Mempengaruhi Umur Pertama Kali Hubungan
Seksual Pada Remaja Menurut Survei RPJMN 2015)**

Proses pengumpulan data sangat rentan mengalami fenomena kekosongan data (*missing data*). Data kosong mengakibatkan estimasi menjadi bias dan standar eror yang semakin besar sehingga dataset menjadi tidak dapat dimanfaatkan kembali. Tujuan penelitian ini adalah membandingkan metode imputasi regresi dan metode MICE pada pembentukan model struktural analisis faktor yang mempengaruhi umur pertama kali melakukan hubungan seksual remaja Indonesia.

Penelitian ini merupakan studi *non reactive* dan menggunakan data Survei RPJMN 2015 yang diambil dari Perwakilan BKKBN Provinsi Jawa Timur. Populasi pada penelitian ini adalah seluruh remaja yang pernah berpacaran dan pernah berhubungan seksual sebelum menikah di Indonesia. Dataset lengkap ini akan dijadikan sebagai data simulasi dengan cara sengaja menghilangkan beberapa data sebanyak 5%, 10%, dan 15% secara random.

Hasil analisis *Friedman Test* menunjukkan bahwa seluruh nilai hasil imputasi tidak berbeda dengan nilai asli ($p < 0,05$). Hasil analisis MSE menunjukkan bahwa MICE mampu menghasilkan nilai MSE lebih kecil daripada imputasi regresi. Hasil analisis GFI dan RMSEA menunjukkan bahwa imputasi regresi memiliki kualitas yang sama baik dengan MICE apabila diaplikasikan pada dataset kekosongan 5% dan 15%. Namun pada kekosongan 10%, MICE lebih baik daripada imputasi regresi.

Kesimpulan bahwa *Multivariate Imputation by Chained Equation* (MICE) lebih baik daripada *Regression Imputation*. Hasil ini diharapkan mampu membantu peneliti lain, khususnya pihak BKKBN, agar tidak membuang unit yang mengandung data kosong. Kelebihan metode MICE adalah dapat diaplikasikan pada data tidak berdistribusi normal dan pada kekosongan hingga 15%.

Kata kunci: Data Kosong, Regression Imputation, MICE, Analisis Jalur