

Duangporn Potawee 2009: The Estimation of Missing Data in Repeated Measurements Using Markov Chain Monte Carlo and Copulas Methods. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Lily Ingsisawang, Ph.D. 99 pages.

The objectives of this research were to study two data imputation methods: Markov Chain Monte Carlo (MCMC) and Copulas, and compare these two methods with the simple form of mean imputation in repeated measures data. The datasets used in this study were simulated by the Monte Carlo technique. Each unit was measured on three occasions under the following conditions: 1) data had multivariate normal distribution with two types of correlation structures: Compound Symmetry (CS) and Autoregressive (AR), 2) the level of inter-correlation among repeated observations are determined as low level (0.3), middle level (0.5) and high level (0.7), 3) the sample sizes used are: 30, 70, and 100 units, and 4) the last measurement was assigned to be missing at random (MAR) by rates of missing data of 5%, 10%, 20% and 30%. These gave rise to a total of 72 possible situations. For the MCMC method, the number of imputations was set to be five. Each defined situation was repeated 1,000 times by SAS program, and the performance of each imputation method was evaluated using mean squared error (MSE). The lower MSE was used to indicate the more effective imputation method.

The results from all situations in the simulation study showed that the Copulas method was the most effective method, yielding the lowest value of MSE. The MCMC method was more effective than the mean imputation when the repeated data had correlation structure as AR. In addition, we applied the above three imputation methods with two datasets in practice: 1) the “Nopparat without Fat, but Healthy” project where the data on waist circumference of participants were repeat measured four times during November 21, 2008 – March 3, 2009; and 2) the data on monthly rainfall in the northern part of Thailand where the Department of Meteorology reported the volume of rainfall by month and by rain station in the north region during June - August 2009. The results also indicated that the Copulas was the most effective method, which was consistent with the simulation data.

---

Student's signature

---

Thesis Advisor's signature