Namthip Phanomthai 2011: A Comparative Study of Missing Data Estimation Methods for Multivariate Data. Master of Science (Statistics), Major Field: Statistics, Department of Statistics. Thesis Advisor: Assistant Professor Boonorm Chomtee, Ph.D. 95 pages.

The purpose of this research is to compare the missing data estimation methods for multivariate data between Expectation Maximization Algorithm (EM) and Regularized Expectation Maximization Algorithm (REM). The criterion of comparison is the mean squares error (MSE). The lower MSE indicates the higher effective estimation method. The datasets used in this study were simulated by the Monte Carlo technique. The comparisons were done under the conditions of independent variables (X) were 3, 4, 5, 7 and dependent variables (Y) were 2 and 3 ; sample sizes (n) are 50, 70 and 100 with the high levels of correlations among independent variables ($\rho = 0.7 - 0.9$) and low levels of correlations among independent variables ($\rho = 0.1 - 0.3$). The dependent variables were assigned to be missing at random (MAR) by 10%, 20% and 30% of missing data rates. These gave rise to a total of 144 possible situations with repeated 1,000 times under each situation. Also, the two real datasets; 1) Stability and forecasting indices group and 2) Moisture group. In upper air observation data during March 1, 2006 - October 31, 2008 from Pimai Radar Station of Bureau of Royal Rainmaking and Agricutural Aviation were used to compare the 2 missing data estimation methods.

The results based on MSE for simulation data showed that in cases of high levels of correlations among independent variables, 3 and 4 of independent variables, sample size 50, at all levels of missing data, EM is a suitable method. The sample sizes 70 and 100 at all levels of missing data, REM is a suitable method. For 5 and 7 of independent variables, at all sample sizes and almost level of missing data, REM is a suitable method. In cases of low levels of correlations among independent variables, 3 and 4 of independent variable, at all sample sizes and almost all level of missing data, EM methods is a suitable method. For 5 and 7 of independent variables, at all sample sizes and almost all levels of missing data, REM is a suitable method. In addition, the result for the two real datasets indicated that REM is a suitable method, which was consistent with that of simulation data.

____________________________ ____________________________ ___ / ___ / ___

Student's signature Thesis Advisor's signature