

## MISSING DATA IMPUTATION USING WEIGHTED OF REGIME SWITCHING MEAN AND REGRESSION

<sup>1</sup>Jumlong Vongprasert and <sup>2</sup>Bhusana Premanode

<sup>1</sup>Faculty of Science, Ubon Rachathani Rajabhat University, Ubon Ratchathani, Thailand

<sup>2</sup>Centre for Bio-inspired Technology, Imperial College London, United Kingdom

Received 2013-11-16; Revised 2014-04-17; Accepted 2014-05-03

### ABSTRACT

Missing data imputation is an important task in cases where it is crucial to use all available data and not discard records with missing values. The purpose of this work were first to develop the Weighted of Regime Switching Mean and Regression (WRSMRI) for missing data estimation and secondly to compare its efficiency of estimation and statistical power of a test under Missing Complete At Random (MCAR) and simple random sampling with another methods, namely; Mean Imputation (MI) Regression Imputation (RI) Regime Switching Mean Imputation (RSMI) Regime Switching Regression Imputation (RSRI) and Average of Regime Switching Mean and Regression Imputation (ARSMRI). By using simulation data, the comparisons were made with the following conditions: (i) Three sample size (100, 200 and 500) (ii) three level of correlation of variables (low, moderate and high) and (iii) four level of percentage of missing data (5, 10, 15 and 20%). The best imputation under MSE and sample correlation estimated were obtained using WRSMRI method, under MAE MAPE power of the test sample mean and variance estimated were obtained using RSRI.

**Keywords:** Missing Data, Imputation, Regime Switching

### 1. INTRODUCTION

Missing data is a common problem that has been found in quantitative research (Heeringa, 2010) albeit there were controlled rigidly in preventive and corrective mechanism (Huisman, 2000). Streiner, (2013) proved that missing of the multivariate random variables by 10% provided analytical errors up to 59%. Estimation of missing data can vigorously improve quality of research in education services (Peng, 2006). For example, in the examination paper impact of missing marking was crucial, in which it could cause errors in both type I and type II (Robitzsch and Rupp, 2009).

On the strong points of the missing data methods, (Sentas and Angelis, 2006) described that in Listwise deletion, cases with missing values for any of the variables are omitted from the analysis. The procedure is quite common in practice because of its simplicity, but when the percentage of missing values is high, it

results in a small complete subset of the initial data sets and therefore in difficulties in constructing a valid cost model. Moreover, the Mean Imputation (MI) method replaces the missing observations of a certain variable with the mean of the observed values in that variable. It is a simple method that generally performs well, especially when valid data are normally distributed. In Regression Imputation (RI) method, the missing values were estimated through the application of multiple regression where the variable with missing data was considered as the dependent one and all other variables as predictors.

On the weak points, Little and Rubin (2002), explained that the values of variance from the LD technique is underestimated. However, Brockmeier *et al.* (2003) tested that the variance from the MI technique is undervalued. Apparently, Little (2005) showed that the RI method conceived the same undervalue, in which it exemplified to a problem of multicollinearity.

**Corresponding Author:** Jumlong Vongprasert, Faculty of Science, Ubon Rachathani Rajabhat University, Ubon Ratchathani, Thailand

This study presents a novel approach in recovery of missing data by employing Weighted of Regime Switching Mean and Regression (WRSMRI). The objectives of this study to compare its efficiency of estimation missing value estimation sample mean, sample variance, sample correlation and power of the test under both Missing Complete At Random (MCAR) and Simple Random Sampling (SRS) with another methods, namely; Mean Imputation (MI) Regression Imputation (RI) Regime Switching Mean Imputation (RSMI) Regime Switching Regression Imputation (RSRI) and Average of Regime Switching Mean and Regression Imputation (ARSMRI).

## 2. MATERIALS AND METHODS

### 2.1. Data Set

In this section, we introduce and describe the data set:

- Three groups of population were simulated data by Monte Carlo technique with three level of correlation of variables (low  $\rho = 0.3$  moderate  $\rho = 0.5$  and high  $\rho = 0.7$ ) (Chaimongkol, 2004; Heeringa, 2010; Little and Rubin, 2002; Viragoontavan, 2000) with size 10,000 units per group
- Sampling methods: We used Simple Random Sampling (SRS) with size with 100, 200 and 500 units. (Chaimongkol, 2004; Viragoontavan, 2000). The data set represented by  $y_1, y_2, \dots, y_n$
- Missing data pattern: We generated missing data using Missing Complete at Random (MCAR) at 5 10 15 and 20% of the sample. (Viragoontavan, 2000). From completed data set we created missing data set by MCAR. The data set split into two groups: Completed data set  $y_1, y_2, \dots, y_r$  and missing data set  $y_{r+1}, y_{r+2}, \dots, y_n$

### 2.2. Methods

In this section, we introduce and describe the methods applied to impute the original incomplete data set and describe the imputation method used based on WRSMRI. The subsequent subsections are organized as follows. First, several general considerations are made to explain how the imputation methods have been implemented. Then, the five imputation techniques applied are described: MI RI RSMI RSRI and ARSMRI. Finally, the WRSMRI method to impute missing value is described together with statistical methods commonly used in methods accuracy evaluation.

### 2.3. Regimes Switching Model

Hamilton (2005) mentioned in a dramatic change in the behavior of a single variable  $y_t$ . Suppose that the typical historical behavior could be described with a first-order autoregression Equation 1:

$$y_t = c_1 + \phi y_{t-1} + \varepsilon_t \tag{1}$$

With  $\varepsilon \sim N(0, \sigma^2)$ , which seemed to adequately describe the observed data for  $t = 1, 2, \dots, t_0$ . Suppose that at date  $t_0$  there was a significant change in the average level of the series, so that we would instead wish to describe the data according to Equation 2:

$$y_t = c_2 + \phi y_{t-1} + \varepsilon_t \tag{2}$$

For  $t = t_0 + 1, t_0 + 2, \dots$ . This fix of changing the value of the intercept from  $c_1$  to  $c_2$  might help the model to get back on track with better forecasts, but it is rather unsatisfactory as a probability law that could have generated the data. We surely would not want to maintain that the change from  $c_1$  to  $c_2$  at date  $t_0$  was a deterministic event that anyone would have been able to predict with certainty looking ahead from date  $t = 1$ . Instead there must have been some imperfectly predictable forces that produced the change. Hence, rather than claim that expression (1) governed the data up to date  $t_0$  and (2) after that date, what we must have in mind is that there is some larger model encompassing them both Equation 3:

$$y_t = c_{s_t} + \phi y_{t-1} + \varepsilon_t \tag{3}$$

where,  $s_t$  is a random variable that, as a result of institutional changes, assume the value  $s_t = 1$  for  $t = 1, 2, \dots, t_0$  and  $s_t = 2$  for  $t = t_0 + 1, t_0 + 2, \dots$ . A complete description of the probability law governing the observed data would then require a probabilistic model of what caused the change from  $s_t = 1$  to  $s_t = 2$ . The simplest such specification is that  $s_t$  is the realization of a two-state Markov chain with Equation 4:

$$\begin{aligned} & \Pr(s_t = j | s_{t-1} = i, s_{t-2} = k, \dots, y_{t-1}, y_{t-2}, \dots) \\ & = \Pr(s_t = j | s_{t-1} = i) = p_{ij} \end{aligned} \tag{4}$$

Assuming that we do not observe  $s_t$  directly, but only infer its operation through the observed behavior of  $y_t$ , the parameters necessary to fully describe the probability law governing  $y_t$  are then the variance of the Gaussian innovation  $\sigma^2$ , the autoregressive coefficient  $\phi$ , the two

intercepts  $c_1$  and  $c_2$  and the two state transition probabilities,  $p_{11}$  and  $p_{22}$ .

The specification in (4) assumes that the probability of a change in regime depends on the past only through the value of the most recent regime, though, as noted below, nothing in the approach described below precludes looking at more general probabilistic specifications. But the simple time-invariant Markov chain (4) seems the natural starting point and is clearly preferable to acting as if the shift from  $c_1$  to  $c_2$  be a deterministic event. Permanence of the shift would be represented by  $p_{22} = 1$ , though the Markov formulation invites the more general possibility that  $p_{22} < 1$ . Certainly in the case of business cycles or financial crises, we know that the situation, though dramatic, is not permanent. Furthermore, if the regime change reflects a fundamental change in monetary or fiscal policy, the prudent assumption would seem to be to allow the possibility for it to change back again, suggesting that  $p_{22} < 1$  is often a more natural formulation for thinking about changes in regime than  $p_{22} = 1$ .

## 2.4. Missing Data Imputation Methods

### 2.4.1. Mean Imputation (MI)

In the general approach to mean imputation, the mean value of each non-missing variable is used to fill in missing values for all observations Equation 5:

$$\hat{y}_j = \frac{1}{r} \sum_{i=1}^r y_i; j = r+1, r+2, \dots, n \quad (5)$$

### 2.5. Regression Imputation (RI)

The completed data set  $(y_1, x_1), (y_2, x_2), \dots, (y_r, x_r)$  used to construct regression equation for impute missing data by Equation 6:

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_j; r+1 \leq j \leq n \quad (6)$$

Where:

$$\begin{aligned} \hat{\beta}_0 &= \bar{y}_r - \hat{\beta}_1 \bar{x}_r \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^r (x_i - \bar{x}_r)(y_i - \bar{y}_r)}{\sum_{i=1}^r (x_i - \bar{x}_r)^2} \\ \bar{y}_r &= \frac{1}{r} \sum_{i=1}^r y_i \quad \bar{x}_r = \frac{1}{r} \sum_{i=1}^r x_i \end{aligned}$$

### 2.6. Regime Switching Mean Imputation (RSMI)

The mean value of each non-missing variable in each group is used to fill in missing values for all observations in group Equation 7:

$$\bar{y}_{j_s_t} = \frac{\sum_{i=1}^{r_{s_t}} y_{i_s_t}}{r_{s_t}}; s_t = 1, 2, \dots, k; j = r+1, r+2, \dots, n \quad (7)$$

### 2.7. Regime Switching Regression Imputation (RSRI)

The completed data set in each group  $(y_{1st}, x_{1st}), (y_{2st}, x_{2st}), \dots, (y_{rst}, x_{rst})$  used to construct regression equation for impute missing data in each group by Equation 8:

$$\hat{y}_{j_s_t} = \hat{\beta}_0 + \hat{\beta}_1 x_{j_s_t}; s_t = 1, 2, \dots, k; j = r+1, r+2, \dots, n \quad (8)$$

Where:

$$\begin{aligned} \hat{\beta}_{0_{s_t}} &= \bar{y}_{rs_t} - \hat{\beta}_{1_{s_t}} \bar{x}_{rs_t} \\ \hat{\beta}_{1_{s_t}} &= \frac{\sum_{i=1}^{rs_t} (x_{i_{s_t}} - \bar{x}_{rs_t})(y_{i_{s_t}} - \bar{y}_{rs_t})}{\sum_{i=1}^{rs_t} (x_{i_{s_t}} - \bar{x}_{rs_t})^2} \\ \bar{y}_{rs_t} &= \frac{1}{rs_t} \sum_{i=1}^{rs_t} y_{i_{s_t}} \quad \bar{x}_{rs_t} = \frac{1}{rs_t} \sum_{i=1}^{rs_t} x_{i_{s_t}} \end{aligned}$$

### 2.8. Average of Regime Switching Mean and Regression Imputation (ARSMRI)

ARSMRI use average of (7) and (8) to impute missing data in each group by Equation 9:

$$\bar{y}'_{j_s_t} = \frac{1}{2} (\bar{y}_{j_s_t} + \hat{y}_{j_s_t}); s_t = 1, 2, \dots, k; j = r+1, r+2, \dots, n \quad (9)$$

### 2.9. Weighted of Regime Switching Mean and Regression (WRSMRI)

WRSMRI use weighted of (7) and (8) to impute missing data in each group by Equation 10:

$$\hat{y}'_{j_s_t} = w_{j_s_t} \bar{y}_{j_s_t} + (1 - w_{j_s_t}) \hat{y}_{j_s_t}; s_t = 1, 2, \dots, k; j = r+1, r+2, \dots, n \quad (10)$$

Where:

$$w_{j_i} = \frac{\text{Var}[\hat{y}_{j_i}]}{\text{Var}[\bar{y}_{j_i}] + \text{Var}[\hat{y}_{j_i}]}$$

### 2.10. Model Evaluation

The accuracy of missing data imputation methods is evaluated by Mean Square Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and power of the test. To evaluate more precisely the difference in prognosis accuracy among the missing data imputation methods, mean square error of sample mean sample variance and sample correlation were evaluated.

## 3. RESULTS

Missing data imputation methods: MI, RI, RSMI, RSRI, ARSMRI and WRSMRI were applied to impute missing data. The goal was to analyses the improvements in accuracy when different algorithms were applied to impute missing data values. **Table 1-3** indicates the average of MSE MAE and MAPE of

imputation methods classified by sample sizes correlation levels and percentage of missing data respectively. **Table 4** indicates the average of MSE MAE and MAPE of imputation methods. **Table 5** indicates the average power of the test of imputation methods. **Table 6-8** indicates the average of MSE MAE and MAPE of sample mean variance and correlation of imputation methods classified by sample sizes correlation levels and percentage of missing data respectively. **Table 9** indicates the average of MSE MAE and MAPE of sample mean variance and correlation of imputation methods.

In terms of MSE WRSMRI outperformed in overall and at sample size 100 and 200, correlation low and high and percentage of missing data 5 15 and 20. RSRI outperformed at sample size 500, correlation moderate and percentage of missing data 10. In terms of MAE and MAPE RSRI outperformed. In terms of power of the test RSRI outperformed.

In terms of sample mean variance and correlation estimated WRSMRI outperformed when estimated sample correlation and RSRI outperformed when estimated sample mean and variance.

**Table 1.** Average MSE MAE MAPE of imputation methods classified by sample sizes.

	Samples size	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
MSE	100	0.7259	0.5417	0.0575	0.0726	0.0450	0.0446
	200	1.5061	1.1131	0.0580	0.0552	0.0456	0.0453
	500	3.6584	2.6557	0.0584	0.0436	0.0461	0.0458
MAE	100	0.1923	0.1648	0.1913	0.1601	0.1688	0.1681
	200	0.1927	0.1638	0.1922	0.1612	0.1699	0.1694
	500	0.1929	0.1631	0.1927	0.1621	0.1707	0.1703
MAPE	100	7.2688	6.2215	7.2298	6.0398	6.3781	6.5818
	200	7.2843	6.1819	7.2650	6.0860	6.4224	6.4021
	500	7.2942	6.1569	7.2861	6.1180	6.4521	6.4345

**Table 2.** Average MSE MAE MAPE of imputation methods classified by correlation levels.

Method	Correlation levels	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
MSE	Low	1.9544	1.7978	0.0568	0.0844	0.0525	0.0525
	Moderate	1.9615	1.4855	0.0584	0.0449	0.0470	0.6074
	High	1.9744	1.0272	0.0587	0.0421	0.0371	0.0365
MAE	Low	0.1906	0.1830	0.1900	0.1800	0.1825	0.1824
	Moderate	0.1933	0.1687	0.1927	0.1657	0.1730	0.1777
	High	0.1940	0.1401	0.1935	0.1377	0.1539	0.1525
MAPE	Low	7.2083	6.9142	7.1870	6.8014	6.9015	6.8960
	Moderate	7.3162	6.3778	7.2932	6.2624	6.5488	6.7234
	High	7.3228	5.2682	7.3007	5.1801	5.8021	5.7515

**Table 3.** Average MSE MAE MAPE of imputation methods classified by percentage of missing data

Method	Percentage of missing data	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
MSE	5	0.7806	0.5694	0.0581	0.0737	0.0457	0.0456
	10	1.5577	1.1363	0.0579	0.0589	0.0455	0.0459
	15	2.4012	1.7654	0.0580	0.0498	0.0455	0.0444
	20	3.1142	2.2762	0.0579	0.0462	0.0454	0.0450
MAE	5	0.1929	0.1641	0.1923	0.1617	0.1702	0.1700
	10	0.1925	0.1638	0.1919	0.1610	0.1697	0.1704
	15	0.1926	0.1639	0.1921	0.1609	0.1697	0.1676
	20	0.1924	0.1639	0.1919	0.1609	0.1696	0.1687
MAPE	5	7.2911	6.1930	7.2691	6.1028	6.4322	6.4234
	10	7.2791	6.1809	7.2558	6.0755	6.4121	6.4401
	15	7.2836	6.1857	7.2620	6.0739	6.4150	6.3331
	20	7.2760	6.1874	7.2544	6.0729	6.4107	6.3773

**Table 4.** Average MSE MAE MAPE of imputation methods

	Methods					
	MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
MSE	1.9634	1.4368	0.058	0.0571	0.0455	0.0453
MAE	0.1926	0.1639	0.192	0.1611	0.1698	0.1692
MAPE	7.2824	6.1868	7.2603	6.0813	6.4175	6.3965

**Table 5.** Average power of the test of imputation methods

		Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
Samples size	100	0.9103	0.9369	0.9105	0.9375	0.9263	0.9265
	200	0.9207	0.9514	0.9207	0.9520	0.9390	0.9394
	500	0.9323	0.9615	0.9323	0.9617	0.9494	0.9497
Correlation levels	Low	0.7713	0.8520	0.7715	0.8534	0.8185	0.8193
	Moderate	0.9920	0.9977	0.9920	0.9978	0.9962	0.9962
	High	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Percentage of missing data	5	0.9332	0.9438	0.9332	0.9438	0.9389	0.9389
	10	0.9258	0.9474	0.9258	0.9476	0.9380	0.9230
	15	0.9157	0.9529	0.9158	0.9536	0.9380	0.9573
	20	0.9097	0.9555	0.9098	0.9566	0.9380	0.9386
Total		0.9212	0.9211	0.9499	0.9504	0.9382	0.9385

**Table 6.** Average MSE of sample mean variance and correlation of imputation methods classified by sample sizes

Method	Samples size	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
Mean	100	0.000063	0.000063	0.000062	0.000045	0.000049	0.000048
	200	0.000031	0.000031	0.000031	0.000022	0.000024	0.000024
	500	0.000012	0.000012	0.000012	0.000009	0.000010	0.000010
Variance	100	0.000074	0.000043	0.000073	0.000041	0.000063	0.000062
	200	0.000069	0.000039	0.000069	0.000038	0.000059	0.000059
	500	0.000066	0.000037	0.000066	0.000036	0.000057	0.000057
Correlation	100	0.002514	0.001474	0.002469	0.001203	0.000694	0.000676
	200	0.001975	0.001053	0.001956	0.000925	0.000351	0.000341
	500	0.001634	0.000806	0.001627	0.000759	0.000144	0.000138

**Table 7.** Average of MSE of sample mean variance and correlation of imputation methods classified by correlation levels

Statistics	Correlation levels	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
Mean	Low	0.000035	0.000044	0.000034	0.000031	0.000032	0.000032
	Moderate	0.000035	0.000037	0.000035	0.000026	0.000028	0.000028
	High	0.000036	0.000026	0.000036	0.000018	0.000023	0.000022
Variance	Low	0.000067	0.000056	0.000067	0.000054	0.000063	0.000063
	Moderate	0.000071	0.000041	0.000070	0.000040	0.000061	0.000061
	High	0.000072	0.000022	0.000071	0.000021	0.000055	0.000053
Correlation	Low	0.001169	0.001235	0.001160	0.000991	0.000575	0.000572
	Moderate	0.001938	0.001249	0.001918	0.001097	0.000398	0.000394
	High	0.003015	0.000849	0.002974	0.000799	0.000216	0.000189

**Table 8.** Average of MSE of sample mean variance and correlation of imputation methods classified by percentage of missing data

Statistics	Percentage of missing data	Methods					
		MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
Mean	5	0.000016	0.000013	0.000016	0.0000120	0.000012	0.000012
	10	0.000028	0.000026	0.000028	0.0000210	0.000022	0.000022
	15	0.000046	0.000046	0.000046	0.0000320	0.000036	0.000035
	20	0.000053	0.000061	0.000052	0.0000380	0.000041	0.000041
Variance	5	0.000011	0.000007	0.000011	0.0000060	0.000009	0.000009
	10	0.000043	0.000026	0.000042	0.0000260	0.000037	0.000037
	15	0.000085	0.000004	0.000084	0.0000440	0.000072	0.000071
	20	0.000146	0.000082	0.000145	0.0000790	0.000125	0.000123
	Correlation	5	0.000423	0.000233	0.000419	0.0002140	0.000162
Correlation	10	.001183	0.000676	0.001170	0.0005970	0.000310	0.000304
	15	0.002646	0.000946	0.002617	0.0011910	0.000513	0.000497
	20	0.004073	0.002232	0.004026	.0001913	0.000624	0.000600

**Table 9.** Average of MSE MAE MAPE of sample mean variance and correlation

Statistics	Methods					
	MI	RI	RSMI	RSRI	ARSMRI	WRSMRI
Mean	0.000035	0.000036	0.000035	0.000025	0.000028	0.000027
Variance	0.000070	0.000040	0.000069	0.000038	0.000060	0.000059
Correlation	0.002041	0.001111	0.002017	0.000962	0.000397	0.000385

#### 4. CONCLUSION

We applied six imputation methods to treat the problem of missing data. We reviewed and provided technical details of the different methods used included MI RI RSMI RSRI ARSMRI and WRSMRI.

As depicted in **Table 1-9**, all imputation methods led to an improvement in prediction accuracy, as measured by MSE MAE MAPE Power of the test MSE for sample mean variance and correlation estimated. The best imputation under MSE and sample correlation estimated were obtained using WRSMRI method, under MAE MAPE power of the test sample mean and variance estimated were obtained using RSRI.

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper.

#### 5. REFERENCES

Brockmeier, L.L., J.D. Kromrey and C.V. Hines, 2003. Nonrandomly missing data in multiple regression analysis: An empirical comparison of ten missing data treatments. Multiple Linear Regress. Viewpoints, 29: 8-29.

- Chaimongkol, W., 2004. Three composite imputation methods for item nonresponse estimation in sample surveys. National Institute of Development Administration.
- Hamilton, J.D., 2005. What's real about the business cycle?, Federal Reserve Bank of St. Louis Review, 87: 435-452.
- Heeringa, S.G., 2010. Multivariate imputation of coarsened survey on household wealth. PhD Thesis, University of Michigan. Dissertation Abstract International.
- Huisman, M., 2000. Imputation of missing item responses: Some simple techniques. Quality Q., 34: 331-351. DOI: 10.1023/A:1004782230065
- Little, R.J. 2005. Missing Data. In: Encyclopedia of Statistics in Behavioral Science, Everitt, B. and D. Howell, (Eds.), Wiley, New York, ISBN-10: 0470860804.
- Little, R.J. and D.B. Rubin, 2002. Statistical Analysis with Missing Data. 2nd Edn., John Wiley and Sons, New York, ISBN: 978-0-471-18386-0, pp: 408.
- Peng, C.Y.J., 2006. Advances in Missing Data Methods and Implications for Educational Research. In: Real Data Analysis, Sawilowsky, S.S. (Ed.), Information Age Publishing Inc., Greenwich, C.T., ISBN-10: 1593115644, pp: 31-78.
- Robitzsch, A. and A.A. Rupp, 2009. Impact of missing data on the detection of differential item functioning: The case of mantel-haenszel and logistic regression analysis. Educ. Psychol. Measur., 69: 18-34. DOI: 10.1177/0013164408318756
- Sentas, P. and L. Angelis, 2006. Categorical missing data imputation for software cost estimation by multinomial logistic regression. J. Syst. Software, 79: 404-414. DOI: 10.1016/j.jss.2005.02.026
- Streiner, D.L., 2013. A Guide for the Statistically Perplexed: Selected Readings for Clinical Researchers. 1st Edn., University of Toronto Press, ISBN-10: 144261353X, pp: 383.
- Viragoontavan, S., 2000. Comparing six missing data methods within the discriminant analysis context: A monte carlo study. Ohio State University.