

Multiple Imputation in Weighting Cell Adjustment

Lee, Joo Young

National Institute of Health, Department of Infectious Disease

Shin, Ming woong & Lee, Jin Hee

Hankuk University of Foreign Study, Department of Statistics

Lee, Sang Eun

Kyonggi University, Department of Applied and Information Statistics

1. Introduction

Multiple imputation is a statistical technique designed to handling the missing data. Here we consider the stratified sampling and apply the discrete case of mutiple imputation method with and without weighting cell adjustment. For discrete case, Bayesian bootstrap(BB) will be used.

2. Multiple Imputaion Method

Let y_{ij} be random sample from i^{th} strata , and each stratum is divided into j^{th} cell. The sample we collected is consisted of responded, $y_{ij}^{(1)}$ and nonresponded data, $y_{ij}^{(0)}$ and can be written as;

$$y_{ij} = \{y_{ij}^{(1)}, y_{ij}^{(0)}\}$$

Here, in this study, we applied the imputation method without and with weighting cell adjustment.

2.1 Without weighting cell adjustment

<i>strata</i>				
1	i	I
y_{11}	y_{i1}	y_{I1}
y_{12}	y_{i2}	y_{I2}
.
.
.
$y_{1(k-1)}$	$y_{i(k-1)_i}$	$y_{I(k-1)_i}$
y_{1k}	y_{ik_i}	y_{Ik_I}

In this case, we apply the imputation method in only considering strata. Let $y_{ik_i}^*$ be the imputed values for $y_{ik_i}^{(0)}$ which is nonresponded sample with k_i^{th} unit in i^{th} strata. Therefore, using BB method, any imputed values for each stratum are from each stratum. Now with imputed values, $y_{ik_i}^*$ $i=1, \dots, I$, we have completed data sets, $y_{i(l)}$ in each i^{th} strata.

And our final result can be obtained by

$$\hat{y} = \sum_{l=1}^M [\sum_{i=1}^I P_i(\bar{y}_{i(l)})] / M$$

where M is number of imputations, $l=1, \dots, M$ and P_i is the sample weight of i^{th} stratum.

2.2 With weighting cell adjustment

For weighting cell adjustment, we divided into several(=J) cells in each stratum. Again using the BB method, imputation values at each cell are obtained from each cell respectively. Then with imputed values, $y_{ijk_j}^*$ $i=1, \dots, I$, $j=1, \dots, J$, we have complete data sets, $[y_{i(l)} | y_{ijk_j}, j = 1, \dots, J]$ in each i^{th} strata, $l=1, \dots, M$.

Then final estimator will be used the same formula as section 2.1. Such as

$$\hat{y} = \sum_{l=1}^M [\sum_{i=1}^I P_i(\bar{y}_{i(l)})] / M$$

where M is number of imputations, $l=1, \dots, M$ and P_i is the sample weight of i^{th} stratum.

However the final estimate will be different due to different completed data set, between with and without weighting cell adjustment. That is $\bar{y}_{i(l)}$ will be based on the adjusted data set.

3. summary

In this presentation, we are expecting the results on with weighting cell adjustment estimator will give the better on MSE value compare with without. Also applying the proper multiple imputation technique will be discussed.

REFERENCES

Rubin, D.B.(1986). Multiple imputation for interval estimation from simple random sample with ignorable nonresponse. *J.A.S.A*, **87**, 366-374.