ISBN: 9789073592421

CPS Poster

Bias corrected imputation method for missing not at random response mechanism using local polynomial regression

Author: Miss Yujin Lee

Coauthors: Key-il Shin

Submission ID: 509

Reference Number: 509

Presentation Image

Bias corrected imputation method for missing not at random response mechanism using local polynomial regression

Yujin Lee[1,2], Key-Il Shin[1]
([1]Hankuk University of Foreign Studies, [2]National Health Insurance Service)

Introduction

Recently, MNAR(missing not at random) nonresponse frequently occurs. Under the MNAR, since the response probability is affected by the study variable, the accurate response probability is not obtained using the observed values of the study variable only.

It is known that using the inaccurate response probability produces bias and so reduces the accuracy of imputation.

In this paper we propose a bias corrected imputation method to improve the accuracy of imputation by removing the bias caused by MNAR.

Assumptions and theoretical bias

To obtain the theoretical bias, we use the following assumptions on response probability model and the super-population model. Using the assumptions, we get the theoretical bias.

[1] Response probability model:

$$p(y_i, x_i) = b_0 + b_1 y_i$$

[2] Super-population model:

$$f_p(y_i|x_i) = m(x_i) + \epsilon_i$$

where $m(x_i)$ is an arbitrary function of the auxiliary variable x_i and $Var(\epsilon_i) = \sigma^2$

[3] Theoretical bias:

$$Bias_i = \frac{b_1}{b_0 + b_1 \mu_i}\sigma^2$$

Where $\mu_i = E(y_i|x_i)$ and $\sigma^2 = Var(y_i|x_i)$

Proposed bias corrected imputation method

Since the usual imputation estimator is biased under MNAR nonresponse, we proposed a bias corrected imputation method. We consider two imputation methods found in R-program: CALIBERrfimpute (mice.impute.rfcont) and predictive mean matching (PMM, mice.impute.pmm). Then the proposed bias corrected imputed estimator is defined by

$$\hat{y}^{BC}_{(k)} = \hat{y}_{(k)} - \widehat{Bias}^{F}_{(k)}$$

where $\hat{y}_{(k)}$ is the imputed value obtained using the explained imputation method and

$$\widehat{Bias}^{F}_{(k)} = \widehat{Bias}_{(k)} \times f^{Bias}, \qquad f^{Bias} = \frac{\hat{B}^{(R)}}{\hat{B}^{(NR)}}$$

$$\hat{B}^{R} = \sum_{i=1}^{r} \widehat{Bias}_i \times \widehat{w}_i^{F}$$

$$\hat{B}^{NR} = \sum_{i=k}^{n-r} \widehat{Bias}_{(k)} \times w$$

Also,

$$\widehat{Bias}_{(k)} = \frac{b_1}{b_0 + b_1 \mu_k}\sigma^2, \qquad \widehat{w}_i^{F} = \frac{w}{\widehat{p}_i}, \qquad w = \frac{N}{n}$$

Here $\widehat{p}_i$ is estimated using the auxiliary variable and local polynomial regression, and then parameters b_0, b_1 are obtained using the regression analysis. Also, μ_i and σ^2 are estimated using local polynomial regression.

Simulation results

We use super-population model:

$$y_i = 10 + 5x_{1i} + 0.2(x_{1i} - 230)^2 + 0.05 sin\left(\frac{2\pi x_{1i}}{100}\right) + \epsilon_i$$

$$\epsilon_i \overset{\text{iid}}{\sim} N(0, 400)$$

Table 1. Non-liner super-population model using CALIBERrfimpute

p_y^{min}	p_y^{min}	r	Bias		ARB		RMSE	
			$\bar{y}^I$	$\bar{y}^{BC}$	$\bar{y}^I$	$\bar{y}^{BC}$	$\bar{y}^I$	$\bar{y}^{BC}$
0.9	0.7	417	-0.0933	-0.0589	0.0258	0.0257	0.717	0.713
0.9	0.5	380	-0.237	-0.160	0.0347	0.0342	0.971	0.954
0.5	0.9	347	0.120	0.0349	0.0407	0.0404	1.120	1.113
0.7	0.9	407	0.0278	-0.00723	0.0273	0.0273	0.749	0.748

Table 2. Non-liner super-population model using PMM

p_y^{min}	p_y^{min}	r	Bias		ARB		RMSE	
			$\bar{y}^I$	$\bar{y}^{BC}$	$\bar{y}^I$	$\bar{y}^{BC}$	$\bar{y}^I$	$\bar{y}^{BC}$
0.9	0.7	417	-0.158	-0.124	0.0717	0.0716	2.001	1.998
0.9	0.5	380	-0.347	-0.269	0.122	0.121	3.390	3.383
0.5	0.9	347	0.197	0.112	0.0919	0.0918	2.607	2.602
0.7	0.9	407	0.0376	0.00262	0.0603	0.0602	1.703	1.702

*Absolute relative bias(ARB) and root mean squared error(RMSE)

Conclusion

In this study, we propose a bias corrected imputation method for MNAR nonresponse. The proposed method can be applied to any existing imputation method. We use a linear response probability model whose bias can be easily calculated, but this proposed method can be applied to any arbitrary unknown super-population model. Simulation studies show that the proposed method can improve the accuracy of the imputation estimator.

Files/Uploads

CPS-Short-Paper-Submission_Yujin Lee_0225

Brief Description

A lot of nonresponse imputation methods have been developed and practically used.

Most imputation methods assume MCAR(missing completely at random) or MAR(missing at random).

Recently, MNAR(missing not at random) nonresponse that are affected by the study variable has occurred frequently, but there are relatively few studies on imputation method for MNAR.

The MNAR nonresponse causes bias and reduces the accuracy of imputation whenever response probability is not properly estimated.

In this study we propose a bias corrected imputation method for MNAR nonresponse under non-informative sampling.

Abstract

Many studies have been conducted to properly handle nonresponse. One method of handling the nonresponse is the nonresponse imputation. A lot of nonresponse imputation methods have been developed and practically used. Most imputation methods assume MCAR(missing completely at random) or MAR(missing at random). Recently, MNAR(missing not at random) nonresponse that are affected by the study variable has occurred frequently, but there are relatively few studies on imputation method for MNAR. The MNAR nonresponse causes bias and reduces the accuracy of imputation whenever response probability is not properly estimated. Practically we do not know both the response probability and the response probability model, and so the best way to impute the missing value is using all information of the available auxiliary variables. In this study we propose a bias corrected imputation method for MNAR nonresponse under non-informative sampling. To estimate the bias, we assume a linear response probability model. The advantage of using a linear response model is that we can obtain the theoretical bias under any super-population model. The obtained bias is applied to some existing imputation methods and we get the bias corrected imputed value. Through simulation studies, we confirm that the suggested bias corrected method gives better results than the existing imputation methods.