

Correction to Min-Ha Lee and Key-II Shin (2022)

In the article “Bias corrected imputation method for non-ignorable non-response” by Min-Ha Lee and Key-II Shin (*The Korean Journal of Applied Statistics*, 2022, Vol. 35, No. 4, p485–499, <https://doi.org/10.5351/KJAS.2023.36.1.correction>), there were errors in section 4.1.4 on page 493.

Section 4.1.4 in Vol. 35, No.4 on p493 was:

4.1.4. 비교 통계량

최종적으로 얻어진 대체값은 다음의 비교통계량, 편향(Bias), 절대상대편향(absolute relative Bias; ARB) 그리고 제곱근 MSE(root mean squared error; RMSE) 결과를 이용하여 성능이 비교되었다. RMSE는 Lin과 Tsai (2020)와 Thomas와 Rajabi (2021)에서 설명한 것처럼 최근 연속형 자료 대체법과 관련된 논문에서 가장 많이 사용된 비교 통계량이다.

$$\begin{aligned} \text{Bias} &= \frac{1}{R} \sum_{k=1}^R \frac{1}{n-r_k} \sum_{i=1}^{n-r_k} (\hat{y}_{k(i)} - y_{k(i)}), \\ \text{ARB} &= \frac{1}{R} \sum_{k=1}^R \frac{1}{n-r_k} \sum_{i=1}^{n-r_k} \frac{|\hat{y}_{k(i)} - y_{k(i)}|}{y_{k(i)}} \times 100, \\ \text{RMSE} &= \sqrt{\frac{1}{R} \sum_{k=1}^R \frac{1}{n-r_k} \sum_{i=1}^{n-r_k} (\hat{y}_{k(i)} - y_{k(i)})^2}, \end{aligned}$$

여기서 $R = 1,000$ 을 사용하였으며 각각의 k 번째 반복마다 새로운 모집단을 생성하여 통계량을 계산하였다. (i) 는 i 번째 결측 자료를 의미하고 따라서 참값인 $y_{k(i)}$ 와 이 자료의 대체 값인 $\hat{y}_{k(i)}$ 의 차이를 계산하였다. 물론 k 번째 반복마다 랜덤으로 결측이 생성되었다. 이는 생성된 특정 모집단의 영향을 줄이기 위함이며 이에 k 번째 반복 모집단의 결측 자료 참값을 $y_{k(i)}$ 로 표시하였다.

The correction is:

4.1.4. 비교 통계량

최종적으로 얻어진 대체값은 다음의 비교통계량, 편향(bias), 절대상대편향(absolute relative bias : ARB) 그리고 제곱근 MSE(root mean squared error : RMSE) 결과를 이용하여 성능이 비교되었다. RMSE는 Lin and Tsai (2020)와 Thomas and Rajabi (2021)에서 설명한 것처럼 최근 연속형 자료 대체법과 관련된 논문에서 가장 많이 사용된 비교 통계량이다.

$$\begin{aligned} \text{Bias} &= \frac{1}{R} \sum_{k=1}^R \frac{1}{n} \sum_{i=1}^n (\hat{y}_{k,i} - y_{k,i}) \\ \text{ARB} &= \frac{1}{R} \sum_{k=1}^R \frac{\left| \frac{1}{n} \sum_{i=1}^n (\hat{y}_{k,i} - y_{k,i}) \right|}{\bar{y}_k} \times 100 \\ \text{RMSE} &= \sqrt{\frac{1}{R} \sum_{k=1}^R \left(\frac{1}{n} \sum_{i=1}^n (\hat{y}_{k,i} - y_{k,i}) \right)^2} \end{aligned}$$

여기서 $R = 1,000$ 을 사용하였으며 각각의 k 번째 반복마다 새로운 모집단을 생성하여 통계량을 계산하였다. $\hat{y}_{k,i}$ 는 대체된 값을 의미하고, $y_{k,i}$ 는 참값을 의미하며 \bar{y}_k 는 k 번째 반복에서 무응답 없이 얻어진 n 개 값의 평균 값으로 참값이다. 또한, 반복마다 랜덤으로 결측이 생성되었다. 이는 생성된 특정 모집단의 영향을 줄이기 위함이다.