Implementing Multiple Imputation and ANCOVA with Rubin's Rule in R: A Comparative Study with SAS

In clinical trials and other research fields, handling missing data effectively is crucial to ensure accurate results. Multiple imputation, a robust statistical method, is commonly employed for this purpose. Traditionally, SAS has been the go-to software for executing multiple imputation and subsequent analysis. However, with the growing popularity of R, an open-source software environment, there is a need to implement these statistical processes in R to offer flexibility and accessibility to the exploratory analysis.

This paper focuses on implementing multiple imputation using both MCMC and monotone regression methods in R, followed by ANCOVA analysis using Rubin’s rule—paralleling established SAS procedures. By leveraging R's extensive statistical libraries, we recreate the multiple imputation process and apply Rubin's rule to analyze the imputed datasets. Through a detailed comparison, we validate the consistency of results between R and SAS, demonstrating that any observed differences are attributed solely to inherent randomness in the imputation process rather than methodological discrepancies.

We utilize dummy data to perform our comparative analysis, illustrating the step-by-step process and ensuring reproducibility. Our findings confirm that R is a viable alternative to SAS for these statistical tasks, offering researchers additional flexibility without compromising accuracy. This work not only contributes to the validation of R as a tool for multiple imputation but also serves as a practical guide for researchers aiming to transition from SAS to R for their statistical analyses.

ABSTRACT:

Accurate handling of missing data is vital in clinical trials and research. Multiple imputation, a robust statistical method, is traditionally executed using SAS software. However, with R's growing popularity as an open-source alternative, implementing these processes in R is increasingly desirable for flexibility and accessibility. This paper details the implementation of multiple imputation and ANCOVA analysis using Rubin's rule in R, replicating established SAS procedures. Leveraging R's statistical libraries, we recreate the multiple imputation process and apply Rubin's rule to imputed datasets. Our comparative analysis, using dummy data, validates the consistency of R's results with those from SAS, attributing differences solely to inherent randomness in the imputation. Our findings confirm R's viability as an alternative to SAS, offering added flexibility without compromising accuracy. This work not only validates R for multiple imputation but also provides a practical guide for researchers transitioning from SAS to R. Key R packages used include "mice" and "norm" for imputation, and "stats" and "emmeans" for ANCOVA.