Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 138
DOI: 10.11159/icsta24.138

Comparison of Methods to Estimate the Average Treatment Effect in Longitudinal Studies When Model Residuals are Not Normally Distributed

Anna Morra, Maya Marintcheva-Petrova, Sonia Amodio
Danone Global Research & Innovation Center
Uppsalalaan 12, 3584 CT, Utrecht
The Netherlands
anna.morra@danone.com; maya.marintcheva-petrova@danone.com; sonia.amodio@danone.commailto:

Extended Abstract

When a traditional longitudinal regression model fits data well, then its residuals should be relatively small, not exhibit any noticeable pattern, have zero mean, constant variance and be normally distributed. Normality of the residuals is important because it justifies the optimized likelihood function and assures that the estimated Least Squares (LS) means are unbiased, and the standard errors are not underestimated. When model's residuals are non-normally distributed, there is the need to investigate possible causes (e.g., model misspecification, presence of outliers and/or influential observations) and to identify potential remedies to draw correct inference (i.e., for estimating standard errors, calculating p-values and confidence intervals). A common strategy is to apply transformations to the outcome variable (e.g. logarithm, squared root) to obtain a distribution closer to normal. However, the drawback of this strategy is less interpretability or even inability to interpret the results on the transformed scale and, depending on the outcome of interest, might not be a viable solution. Moreover, in some cases data transformations are not successful in improving the fit of the model, and, therefore, other approaches need to be considered. A further complication in the analysis of data from longitudinal trials is the handling of unavoidable missing values due to premature study discontinuation or skipped visits.

We explore the combination of multiple imputations and robust regression [1] to estimate the Average Treatment Effect (ATE) at various timepoints, in longitudinal studies when the traditional mixed effects model for repeated measurements results in non-normally distributed residuals and compare the results in terms of bias and inflation of standard errors with other available approaches e.g., weighted Generalized Estimating Equations (WGEE), bootstrap techniques and most recent published methodologies [2]. We base our investigations on simulated data that emulates the characteristics of a randomized parallel clinical trial with repeated measures of infant's anthropometrics (e.g., weight) over the first 4 months of life. In this context, we simulate scenarios related to different violations of distributional assumptions, including heavy tails in symmetric distribution and skewed distributions [3], and give recommendations about the most suitable approach in different scenarios.

References

[1] D. V. Mehrotra, X. Li, J. Liu and K. Lu, "Analysis of Longitudinal Clinical Trials with Missing Data Using Multiple Imputation in Conjunction with Robust Regression", *Biometrics,* vol. 68, no. 4, pp. 1250-1259, 2012.
[2] S. Liu, Y. Zhang, G. T. Golm, G. Liu and S. Yang, "Robust analyzes for longitudinal clinical trials with missing and non-normal continuous outcomes", *Statistical Theory and Related Fields,* vol. 8, no. 1, pp. 1-14, 2023.
[3] H. Schielzeth, N. J. Dingemanse, S. Nakagawa, D. F. Westneat, H. Allegue, C. Teplitsky, D. Réale, N. A. Dochtermann, L. Z. Garamszegi, Y. G. Araya-Ajoy. "Robustness of linear mixed-effects models to violations of distributional assumptions", *Methods Ecol Evol*, vol 11, no. 9, pp. 1141-1152, 2020.