[BIOINFO 2023]

Two-Part Mixed-Effects Model for Batch Effect Correction in Microbiome Data

Seungrin Yang[1], Taesung Park[1,2*]

[1,2] *Interdisciplinary Program in Bioinformatics, Seoul National University*
[2] *Department of Statistics, Seoul National University*
**Corresponding author: tspark@stat.snu.ac.kr*

In the field of microbiome research, batch effects pose a significant challenge to the accurate interpretation of data. This study introduces a novel two-part model for batch effect correction in microbiome data analysis, which combines a modified mixed logistic regression model and a negative binomial regression model. The mixed logistic regression model, which models the presence or absence of taxa, incorporates batch ID as a random effect variable and clinical factors as fixed effect variables. The negative binomial regression model, which models the count of taxa. Notably, the mixed logistic regression model is modified to generate a "batch-free" model by subtracting the random effect term from the expected logit value. The performance of the proposed model was evaluated using PERMANOVA R^2 values under both Bray-Curtis and Aitchison dissimilarities, and compared with existing methods. The results demonstrated that the proposed model effectively reduces the batch effect and provides a robust solution for batch effect correction in microbiome data. This study contributes to the advancement of microbiome data analysis by offering a novel approach to address the critical issue of batch effects in microbiome data.