Simple Two Level Monte Carlo Simulation

Simulation in R Final Project

This project focused on running a simple multi-level model with two levels and a treatment condition. There were two conditions for this simulation: variation at the level two units and varying the ICC. The variation was set using Cohen’s d and was set at either 0.1 or 0.5 The ICC was chosen to be either 0.10 or 0.25. In total, there were 4 different conditions shown in Table 1:

**Table 1:** Conditions

|  |  |  |
| --- | --- | --- |
|  | ICC | Cohen’s D |
| 1 | 0.10 | 0.10 |
| 2 | 0.25 | 0.10 |
| 3 | 0.10 | 0.50 |
| 4 | 0.25 | 0.50 |

For this simulation, there were 5 other baseline parameters. There were 1000 level 2 clusters with 100 individuals per cluster bringing the total number of participants to 100000 per dataset. The total variance of the outcome variable was set to be 100 while the mean was set to be 25. The full set of parameters are shown in Table 2:

**Table 2 :** List of Parameters

|  |  |
| --- | --- |
| Parameter | Value |
| Number of Clusters (J) | 1000 |
| Individuals per cluster (nJ) | 100 |
| ICC | [0.10, 0.25] |
| Total Variance of Outcome | 100 |
| Mean of Outcome | 25 |
| Cohen’s D | [0.10, 0.50] |

The data generating model was based on Equation 1 below:

Generating the model was done through a function that determined that set up each part of the dataset. First the ids for both levels were set through simple vector creation. The treatment indicator was set at the second level by giving half of the level 2 ids a value of 1 and the other half a value of 0. The variances at each level was then determined. The variance at level is shown in Equation 2 while the total variance at level 2 is shown in Equation 3

Gamma 1 (γ­1) was determined based on the Cohen’s D value and the standard deviation of the outcome and is shown in Equation 4

The portion of the outcome variable from within effects was determined using a normal distribution displayed in equation 5.

The Level 2 Variance Explained was determined using the ICC and the gamma from equation (4) and is displayed in equation (6). The residual variance was determined by subtracting the level 2 variance explained from the total level 2 variance (7).

The between outcome effect was determined by using Equation 8. The treatment variable was multiplied by γ­1­ and then added to a normal distribution to account for the level 2. This was then added to the within portion of the outcome to determine the total value of the outcome variable (9).

This was all put into a single function that was run for the 4 different parameter conditions. The function was then replicated to produce a list of datasets that was then evaluated for type 1 error rates through coverage. The evaluation was done using three methods: Standard linear modeling, multilevel modeling, and robust cluster standard error. Each data set was then evaluated to determine the percentage that the mean difference of treatment effect was within a 95% confidence interval of the population parameter. The results of the condition and the percentage that the gamma was within the 95% confidence is shown in Table 3:

**Table 3**: Evaluation Criteria

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Standard LM | Multi-Level Model | Robust Standard Error |
| 1 | 92.8% | 94.8% | 92.8% |
| 2 | 97.4% | 97.4% | 97.4% |
| 3 | 93.8% | 95.2% | 94.8% |
| 4 | 89.8% | 94.0% | 94.0% |

Here, the MLM and Robust Standard error were fairly similar (except for Condition 2) while the Standard LM is farther away (especially in condition 4).