7. Summary and Conclusions

The purpose of this study was to explore the associations between sample sizes at different levels of clustered data and the sampling precision of the results derived from hierarchical linear models (HLM). As outcomes, we provided graphs and equation tables that show the connection between the two concepts.

Our research was strictly oriented toward the needs of researchers wanting to apply HLM analyses to data collected in large-scale educational surveys (LSA), or survey designers with similar interests. We considered the specifics of such datasets and used a Monte Carlo simulation study in order to explore various population and sample conditions. In particular, the explored settings varied in terms of the following:

- Sample sizes of and within clusters;
- Intraclass correlation coefficients;
- Covariance distribution; and
- Weight status.

We explored, for all settings, four different hierarchical linear models with increasing complexity. We used the coefficient of variation, displayed as a percentage, to measure sampling precision.

On average, over all explored settings and models, the parameters γ_{00} (mean of random intercepts) and ε (residual variance) could be measured with the highest sampling precision levels. On the contrary, the parameter γ_{01} (slope of random intercepts) was the parameter that was measured with the poorest precision. This was particularly the case when the covariance distribution between the outcome and the explanatory variable was stronger at the within-cluster level.

As we expected, the coefficients of variation of all explored parameters decreased when sample size increased. The dependency between sample size and coefficient of variation could always be described by a quadratic curve progression, within the explored setting, such that increasing sample size decreased the diminishing effect on the coefficient of variation. This general observation was affected neither by the intraclass correlation coefficients, the weight status, or the covariance distribution, nor by the complexity of the explored model. The magnitude of this decrease, and whether the effect was more pronounced with sample size increases on one or the other hierarchical level, could depend, however, on all these factors and was different for the explored model parameters.

In conclusion, the results showed that the required sample sizes depended heavily on the parameter of interest. In particular, sample size requirements differed widely for the estimation of fixed-model parameters and the estimation of variances. In agreement with the literature, it appears that increasing the number of sampled clusters rather than the cluster sample size is more effective if the research interest concerns macro-level regression coefficients. If the focus is on variance estimates, however, the level on which the sample size is increased appears to be of less importance.

It is worthwhile noting that the reduction in the coefficient of variation of the variance of random intercepts (parameter U_0) seemed to become notably larger as we stepped from 5 to 10 sampled units per cluster, especially for low ICCs. In fact, the gain in precision was not that much larger when, for example, doubling sample sizes at Level 1 than when doubling sample sizes at Level 2. This finding could have particular relevance with respect to cost considerations.

The intraclass correlation coefficient had no influence on the coefficient of variation of the residual variance, while the coefficients of variation of the parameters γ_{00} (mean of random intercepts) and γ_{01} (slope of random intercepts) increased with larger ICC values. This effect diminished with larger within-cluster sample sizes. Exploring the impact of the ICC on the coefficients of variation of the remaining considered parameters—U_0 (variance of random intercepts), γ_{10} (mean of random slopes), and γ_1(fixed slope)—produced the inverse effect: a decrease in the coefficients of variation as the ICC levels increased.

As a new contribution to this research area, we considered two cases of covariance distribution. We found that the effect of the covariance distribution on the coefficients of variation of the parameters β_1 (fixed slope) and particularly γ_{01} (slope of random intercepts) was—at least within the limitations/conditions of this research—even more pronounced than the effect of varying sample sizes. We noted no effect on the coefficients of variation of any explored parameter, other than the fixed slope, in Models 1 and 2. We also observed no effect on parameter ε in any model.

Compared to the second considered case of covariance distribution (10 at within-group and 20 at between-group level) for the parameters γ_{00} and γ_{01}, the first considered case of covariance distribution (20 at within-group and 10 at between-group level) was connected to higher coefficients of variation. The differences in the coefficients of variation were extreme for parameter γ_{01}, particularly when explanatory variables on both levels were introduced to the model. For the parameters U_0, β_1, and γ_{10}, the coefficients of variation were higher if the covariance distribution was 20 between groups and 10 within, rather than the other way around.

Weights, which have to be applied to allow unbiased estimates in LSA, enlarged the coefficients of variation of all explored parameters consistently by a factor of approximately 1.1. Preliminary evaluations of the findings with real data showed that this factor held true only if the actual Level 2 weights followed a Poisson distribution. This was the case if the implemented sample design fulfilled particular conditions, detailed in the respective chapter.

Model complexity had an influence on the coefficients of variation of all observed parameters except for the residual variance. The influence varied with the parameter of interest as well as with the considered case of covariance distribution. The coefficients of variation of parameter γ_{00} were smaller in the models that included a macro-level explanatory variable. The effect was more pronounced in the case where the covariance was stronger between clusters.

For parameters U_0 and β_1, however, the precision diminished with increasing model complexity, but the differences remained marginal as long as the covariance was strong at the within-cluster level. For these two parameters, the effect decreased with increasing ICC levels. The coefficient of variation of parameter γ_{01} also increased with model complexity, but the effect here was clearly more pronounced when the covariance was stronger within clusters.

We end by emphasizing that all findings can be deemed valid only within the explored ranges of sampling and population settings. The degree of generalizability of the results to other settings and conditions will be the subject of further research.