# Between-group and within group effects and intra-class correlation

**Juha Lappi**[1]

[1] *University of Eastern Finland, Finland, juha.lappi.sjk@gmail.com*

Two aspects of between-group and within group effects in grouped data are discussed, when a constant intra-class correlation of the residual errors is assumed in a linear model. A positive intra-class correlation can be desribed with a random effects model. First, it is shown how the dependency of variances of GLS estimates on the values of the regressor variables can be described in terms of the intra-class correlations of the regressor variables. The effect of an regressor $x$ in the diagonal of moment matrix can be decribed as

$$\mathbf{X}'\mathbf{W}\mathbf{X}_{xx} = \sum_{i=1}^{K}\sum_{j=1}^{n}\left[\left(1\left(n-1\right)\rho_e\right)\left(x_{ij}-\bar{x}_i\right)^2 + \left(1-\rho_e\right)\left(\bar{x}_i-\bar{x}\right)^2\right]$$

That is, the deviation of deviation of $x$ from the group mean gets different weight than the group mean. When the intra-class correlation $\rho$ of the residuals is large, then $x_{ij}-\bar{x}_i$ has large weight. Large intra-class correlation of $x$ indicates small variation of $x_{ij}-\bar{x}_i$. Thus, then both correlations are large, then also the variance of the the GLS estimator is large. Derivations in terms of $x_{ij}-\bar{x}_i$and $\bar{x}_i$ explains nicely the relations between OLSE, GLSE and BLUE. In the case of singular covariance matrix of residual errors, this presentation also leads to a derivation a new BLUE which combines OLS and GLS estimation principles in the same estimator.

Using regressor $x$ in a linear model assumes that $x_{ij}-\bar{x}_i$ and $\bar{x}_i$ have the same coefficient. When $x_{ij}-\bar{x}_i$ and $\bar{x}_i$ are used as different regressors, the estimation of the model has special features. For instance, adding $x_{ij}-\bar{x}_i$ to the model increases the estimated variance of the group effect of the residual error. It is natural to assume that the whole group mean of $x$ and the deviation from the whole group mean have an effect. When the whole group is not measured, this leads to bias problems as in other cases where regressors are measured with error. The bias is corrected in [1] using a random effects assuptions also for $x$. This topic is discussed there in a greater detail. It is suggested that negative intra-class correlations should be given more attention.

## References

[1] Mehtätalo, L. and Lappi, J. (2020)*Biometry for forestry and environmental data with examples in R*, CRC Press,Boca Raton