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 ρ 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, when 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

Nonparametric graphical tests of significance for the functional general linear model with application to forestry data

Mari Myllymäki[1]

[1] *Natural Resources Institute Finland (Luke), Finland, mari.myllymaki@luke.fi*

Permutation methods are commonly used to test significance of regressors of interest in general linear models (GLMs) for functional data sets, as they rely on mild assumptions. Permutation inference for GLMs typically consists of three steps: choosing a relevant test statistic, computing pointwise permutation tests and applying a multiple testing correction. In this talk, I will present test statistics that together with the global envelope tests applied as the multiple testing correction allow for useful graphical interpretation of the test results [1, 2, 3]. As such, the tests are able to find not only if the factor of interest is significant but also which functional domain is responsible for the potential rejection, and between which groups of a categorical factor the differences occur. I will discuss the use of the graphical tests for examining the influence of species and other forest stand attributes on the vertical distribution of aerial light detection and ranging (LiDAR) returns [4]: Tree species have different shapes and forest stand dynamics. LiDAR is a technique that can measure the three dimensional position of reflective material by sending laser pulses through the canopy. We examined the influence of species, crown closure, and age on the vertical distribution of aerial LiDAR returns of regular, even-aged forest stands in Quebec, Canada.

References

[1] Mrkvička, T., Myllymäki, M., Jilek, M. and Hahn, U. (2020). A one-way ANOVA test for functional data with graphical interpretation. *Kybernetika* **53**(3), 432 – 458. https://doi.org/10.14736/kyb-2020-3-0432

[2] Mrkvička, T., Roskovec, T. and Rost, M. (2019). A nonparametric graphical tests of significance in functional GLM. *Methodology and Computing in Applied Probability.* https://doi.org/10.1007/s11009-019-09756-y

[3] Myllymäki, M. and Mrkvička, T. (2020). GET: Global envelopes in R. arXiv:1911.06583 [stat.ME]

[4] Racine, E., Coops, N.C., Bégin, J. and Myllymäki, M. (2021). Species, crown closure, and age as determinants of the vertical distribution of airborne LiDAR returns. arXiv:2104.05057 [q-bio.QM]

Finding hidden trees in remote sensing of forests by using stochastic geometry, sequential spatial point processes and the HT-like estimator

Lauri Mehtätalo[1], Kasper Kansanen[2] and Adil Yazigi[2]

[1] *Natural resources institute Finland (LUKE), Joensuu, lauri.mehtatalo@luke.fi*
[2] *University of Eastern Finland, Joensuu*

Remote sensing is increasingly used in collecting data for forest inventories. Especially, aerial and terrestrial laser scanners are used to collect three-dimensional information about the forest. In aerial laser scanning, a laser scanner is installed to airplane or a drone and the device is used to collect information about the forest canopy below. Individual tree detection can further be done using the laser point cloud to extract individual tree crowns and estimate the tree heights and tree crown size and shape. In terrestrial laser scanning, a scanner is placed on a tripod and the surrounding forest is scanned for measurements of tree stems. Individual tree stems can be detected from the point cloud and used for estimation of tree DBH and other tree characteristics.

In both abovementioned cases, the detected trees can be ordered according to their shortest distance to the sensor. Because of this hierarchical ordering of trees, trees that are "earlier" in the hierarchy can cause "latter" trees to remain undetected because they are located in the hidden area or sector formed by the earlier trees. These undetected trees cause problems in inventories where the population totals per area-unit are of interest. We discuss a recently introduced Horvitz-Thompson-like estimator for the population totals in this setting, which is unbiased in certain conditions if the tree locations follow the complete spatial randomness [1, 2]. We also present approaches to take into account the spatial pattern of tree locations in the approach by using a ordered spatial point process model [3].

References

[1] Kansanen, K., P. Packalen, M. Maltamo, and L. Mehtätalo. (2021). Horvitz-Thompson-like estimation 521 with distance-based detection probabilities for circular plot sampling of forests. *Biometrics* early view. https://onlinelibrary.wiley.com/doi/abs/10.1111/biom.13312.

[2] Kansanen, K., J. Vauhkonen, T. Lähivaara, and L. Mehtätalo. (2016). Stand density estimators based on individual tree detection and stochastic geometry. *Canadian Journal of Forest Research* 46(11):1359http://dx.doi.org/10.1139/cjfr-2016-0181.

[3] Yazigi, A., A. Penttinen, A.K. Ylitalo, M. Maltamo, P. Packalen, and L. Mehtätalo. (2019). Sequential spatial point process models for spatio-temporal point processes: A self-interactive model with application to forest tree data. *arXiv preprint*