Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 124
DOI: 10.11159/icsta25.124

Modelling of the Short-term Heart Rate Variability Measures: a Focus on the Transformation

Svitlana Shvydka and Mária Ždímalová
Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering
Slovak University of Technology in Bratislava
Radlinského 11, 810 05 Bratislava, Slovak Republic
svetlana.shvydka@gmail.com, svitlana.shvydka@stuba.sk; maria.zdimalova@stuba.sk

Extended Abstract

Medical data often includes nonnegative continuous outcomes that are skewed and contain zero observations. A square root transformation works well with such datasets. When there are no zeros, a logarithmic transformation may be more appropriate. Transformations are aimed to create a more symmetric distribution, which simplifies subsequent statistical analysis [1]. With applying the generalized linear and additive models the transforming data largely unnecessary, yet the logarithmic and square root transformations are still widespread to model the medical data.

To illustrate this point the empirical data from Danuta Makowiec [2] was used. From the set of standard heart rate variability (HRV) indices, the following measures were selected: pNN50 - the percentage of adjacent normal-to-normal intervals that differ from each other by more than 50 ms, RMSSD – the root mean square of successive differences between normal heartbeat(s), PSS – the probability of short segments. Among the 42 heart transplant (HTX) patients, 12 (29%) patients had a left ventricle of normal geometry (group 1), 22 (52%) patients had concentrically remodeled (group 2), and 8 (19%) patients demonstrated hypertrophic remodeled (group 3) one year after HTX.

Among the generalized additive models (GAMs) for untransformed and transformed (log or square root) data of the parameters RMSSD, PSS and pNN50 only GAMs for index pNN50 showed the statistically significantly difference between groups of patients. The following GAMs with Gaussian distribution were estimated. M1: pNN50~groups; M2: sqrt(pNN50)~groups; M3: $\log_{10}$(pNN50+0.001)~groups; M4: $\log_{10}$(pNN50+0.01) ~groups; M5: $\log_{10}$(pNN50+0.1) ~groups. The untransformed and transformed (sqrt(pNN50), $\log_{10}$(pNN50+0.1)) data showed positively skewed distributions. The Shapiro–Wilk test fails for all transformation scenarios, however, the distributions reveal a near-symmetric shape for $\log_{10}$(pNN50+0.001) and $\log_{10}$(pNN50+0.01) values. In terms of statistical significance, all models showed the consistent results (M1: edf = 1.66, p=0.04, R^2(adj)=0.11, deviance explained =13.9%, Akaike Information Criterion (AIC)=-32; M2: edf = 1.72, p=0.03, R^2(adj)=0.12, deviance explained =15.7%, AIC=-20; M3: edf = 1.69, p=0.03, R^2(adj)=0.12, deviance explained =15.2%, AIC=-177; M4: edf = 1.77, p=0.03, R^2(adj)=0.13, deviance explained =16.2%, AIC=-161; M5: edf = 1.70, p=0.048, R^2(adj)=0.10, deviance explained =13.3%, AIC=-106; where edf is estimated degrees of freedom, p-value is used in determination of statistical significance, R^2(adj) is adjusted coefficient of determination). According to the GAMs, the patients of the group 3 have a higher probability of having higher values of the index pNN50 than the patients of the groups 1 and 2. The quantile-quantile plots and histograms of residuals were checked. The model M4 was the best among all the fitted models.

Variable transformation is a powerful tool to deal with skewed data. However, the transformed data does not always represent a better approximation to the normal distribution. Additionally, the transformation can complicate interpreting the results on the original scale, especially when complex or multiple transformations are involved, making back-transformation a challenging process.

Acknowledgements

Svitlana Shvydka is funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V01-00029. The work of Svitlana Shvydka and Mária Ždímalová are partially supported by the grant VEGA 1/0036/23. Svitlana Shvydka thanks Nicolas Desreumaux from the Observatory Oceanology de Banyuls-sur-Mer FR 3724, Sorbonne University, for advice and comments on data analysis and modelling.

References

[1] A. F. Zuur, E. N. Ieno, “A protocol for conducting and presenting results of regression-type analyses,” *Methods Ecol. Evol*., vol. 7, no. 6, pp. 636-645, 2016.

[2] D. Makowiec, J. Wdowczyk, M. Gruchała, “Heart rate variability by dynamical patterns in windows of holter electrocardiograms: a method to discern left ventricular hypertrophy in heart transplant patients shortly after the transplant,” *Biomedinformatics*, no. 3, pp. 220-251, 2023.