

Human growth data analysis and statistics – the 5th Gülpe International Student Summer School

Detlef Groth¹  • Christiane Scheffler²  • Michael Hermanussen³ 

¹ University of Potsdam, Institute of Biochemistry and Biology, Bioinformatics, 14476 Potsdam, Germany.

² University of Potsdam, Human Biology, 14469 Potsdam, Germany.

³ Aschauhof 3, 24340 Eckernförde – Altenhof, Germany.

Citation:

Groth, D./ Scheffler, C./Hermanussen, M. (2023). Human growth data analysis and statistics – the 5th Gülpe International Student Summer School, Human Biology and Public Health 1. <https://doi.org/10.52905/hbph2023.1.70>.

Received: 2023-05-18

Accepted: 2023-05-25

Published: 2023-07-21

Copyright:

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Correspondence to:

Detlef Groth
email: detlef.groth@uni-potsdam.de

Abstract

The Summer School in Gülpe (Ecological Station of the University of Potsdam) offers an exceptional learning opportunity for students to apply their knowledge and skills to real-world problems. With the guidance of experienced human biologists, statisticians, and programmers, students have the unique chance to analyze their own data and gain valuable insights. This interdisciplinary setting not only bridges different research areas but also leads to highly valuable outputs. The progress of students within just a few days is truly remarkable, especially when they are motivated and receive immediate feedback on their questions, problems, and results. The Summer School covers a wide range of topics, with this year's focus mainly on two areas: understanding the impact of socioeconomic and physiological factors on human development and mastering statistical techniques for analyzing data such as changepoint analysis and the St. Nicolas House Analysis (SNHA) to visualize interacting variables. The latter technique, born out of the Summer School's emphasis on gaining comprehensive data insights and understanding major relationships, has proven to be a valuable tool for researchers in the field. The articles in this special issue demonstrate that the Summer School in Gülpe stands as a testament to the power of practical learning and collaboration. Students who attend not only gain hands-on experience but also benefit from the expertise of professionals and the opportunity to engage with peers from diverse disciplines.

Take-home message for students Learning by analyzing your own data is essential to understand statistical concepts.

“We are what we repeatedly do. Excellence, then, is not an act but a habit” wrote Will Durant (1926) in “The Story of Philosophy” summarizing Aristotle’s thoughts in the “Nicomachean Ethics” (Durant 1991).

Already the Greek philosopher Aristotle knew that learning through practical exercise is the crucial component of education. By practical exercise individuals are enabled to apply their theoretical knowledge to real-world situations. The goal of the Gülpe Summer School is to provide hands-on learning for students with their own datasets, therefore providing as much motivation as well as pleasure due to the wonderful landscape of northwestern Brandenburg.

One of the areas where students have often a theoretical foundation but lack practice is statistics. Here, human biologists often face challenges in understanding statistical principles due to divergent terminology that differs from their thinking process. However, our Summer Schools successfully bridge this gap by introducing statistical techniques tailored to our students’ needs. Surprisingly, even the introduction of tools with which the students have never worked before, like the statistical programming language R (R Core Team 2022), has been well-received, as our students are highly motivated to analyze their own data. This motivation, coupled with the combination of engaging lectures and immediate application of newly acquired knowledge to their own data, has proven to be highly effective. This approach offers a solid foundation in statistical principles and techniques allowing the students to navigate through the vast ocean of statistical terms, tests, and plots. To assist them further, we provide basic guidelines, as exemplified in Table 1, which offer an initial orientation and serve as valuable references to apply techniques and methods in descriptive and inferential statistics.

We assume here that the variables are either of categorical or continuous scale type. Depending on the scale type of the variable or the number of groups, e.g., for a categorical variable, things like the statistics used, the test chosen, or the effect size should be selected. For instance, to look for differences in means for normally distributed data, where the categorical variable has more than two groups, we would use an ANOVA, while for two groups, we would use a t-test.

Such guidelines, as presented in Table 1, provided in the short introductory lectures, offer new insights to students, enabling them to finally make sense of concepts they may have previously encountered already during their study times. In the Summer School, the students then finally apply these concepts to their own data. However, given such rules and guidelines, the final stage of statistical analysis may seem easy to accomplish, but now frustration on the student’s side often arises during the data preparation phase, which often feels like a nightmare. Despite the lecturer’s warning that data preparation can be the most challenging part, students may have difficulties getting their data into the analytical pipeline. However, it is important to explain that successful data preparation is crucial for accurate and meaningful statistical analysis.

Therefore, despite initial failures, students must develop resilience to ensure that their data is ready for analysis at the end.

Another type of surprise that often occurs is when students realize they cannot simply hand over their data to the lecturers for analysis. Instead, the lecturers empower the students to conduct their own analyses, fostering their independence and self-reliance. This realization is particularly impactful as students learn to take ownership of all aspects of data analysis, including data preparation, descriptive analysis, and inferential analysis. By developing these

Table 1 Main descriptive and inferential methods and measures for bivariate data statistics.

	cat ~ cat	norm ~ cat	no-norm ~ cat	norm ~ norm	norm ~ no-norm
center	modus	mean, modus	median, modus	mean, mean	mean, median
spread	proportions	sd (cv), prop.	skewness, modus	sd (cv)	sd, skewness
plot	assocplot	boxplot	boxplot (of log normalized data)	xyplot	xyplot (of untransformed vs log normalized data)
test	chisq	t-test, ANOVA	wilcox, kruskal	Pearson	Spearman
effect-size	Cohen's w	Cohen's d, Eta-squared	Wilcoxon r, Epsilon-squared	Pearson's r	Spearman's rho

skills, they become less reliant on external statisticians and their limited availability. So finally, this offers them greater control over their own research and enhances the quality of their findings. Standing on their own feet, students will develop a sense of confidence and competence that will serve them well throughout their academic and professional career.

Two students focused on statistical algorithms. Nikolaos Gasparatos (Gasparatos et al. 2023) delved into the problem of detecting change points in time series data. Gasparatos examined the identification of points where steadily increasing values of a variable experience a dramatic increase within a short period of time. This type of analysis, known as changepoint analysis, plays a critical role in identifying significant shifts in data trends. Gasparatos utilized synthetic data to evaluate and refine the method and finally would like to apply the developed model to real biological data, particularly to detect so-called “mini growth spurts”. Mini growth spurts are rapid accelerations of growth velocity within a few days during children’s development (Hermanussen et al. 1988). The outcome of his analysis primarily indicated that the suggested algorithm works if the growth spurts are sufficiently large. However, at the end it became evident that the present model needs verification and fur-

ther extension to be suitable for real-world scenarios. Therefore, the ultimate question: “Is my developed model suitable to be used for the prediction of mini growth spurts in children?” still remains to be resolved.

Tim Hake et al. (Hake et al. 2023) also focused on a more theoretical aspect of data analysis. His work involved translating and extending an existing algorithm for the St. Nicolas House Analysis (SNHA) (Groth et al. 2019) from R to Python. The SNHA algorithm itself can be considered as a true “Summer School child”, it was developed to solve the old nightmare of “everything is correlated with everything” and to provide an initial understanding of the most essential associations in the students’ data. In the early Summer Schools, Principal Component Analysis (PCA) was primarily used for this purpose. However, it has been shown that SNHA’s network approach and the simplicity of the algorithm provide results that are much easier to interpret. Recent publications and the snha R package (Groth 2023) show that the SNHA algorithm has outgrown its infancy stage. Since not every statistician is also an R user, it is important that users of other programming languages widely used in statistics, such as Python, can also use this algorithm. Hake not only ported the algorithm to Python but also evaluated enhancements to the algorithm by incorpo-

rating bootstrap techniques to facilitate the detection of associations between variables in dense network structures. In networks consisting of numerous variable associations, the chain of associations becomes more intricate. The findings presented in Hake's paper demonstrate that the utilization of resampling methods can improve the detection of first-class associations between variables even in dense network structures. The bootstrapping technique offers valuable statistical insights into the quality of the identified edges.

With the article of Malambo and Musalek (Malambo and Musalek 2023) we delve into the realm of real-world data analysis. Malambo explored the association between intelligence quotient (IQ), and body mass index (BMI) and body height in preschool children. Similar to what many other students experienced in the recent years, his analysis started with a St. Nicolas House Analysis. Yet, the initial approach revealed a surprising lack of associations among these variables. In contrast to previous studies involving older children and adults (Corley et al. 2010), Malambo failed to find significant associations between anthropometric measures and cognitive data. This finding may be less surprising given the fact that early childhood is more an open window of opportunity than a definitive endpoint in the child's development.

Antonia Rösler (Rösler et al. 2023) assessed whether forced migration after the Second World War in Poland resulted in growth impairment among children. The analysis focused on a data set comprising more than 2028 individuals and aimed to identify significant differences in the main anthropometric measures such as height or weight. However, the analysis, finally selecting 898 girls from the dataset, revealed no such differences indicating a remarkable resilience of children's growth even against serious

environmental stress. These results highlight the adaptive nature of child growth and their ability to withstand difficult environmental conditions.

The manuscripts included in this issue are the culmination of the students' and supervisors' efforts to perform collaborative research, utilizing synthetic and original data. During the Summer School, these datasets and analytical approaches underwent rigorous evaluation and discussion. The manuscripts highlight the collective dedication and the collaborative spirit exhibited at the Summer School. It is truly remarkable to witness the rapid progress that the students can achieve within just a few days of intense and fruitful work, fostered by the inspiring atmosphere at the magnificent Ecological field station in Gülpe.

Acknowledgements

The Summer School was financially supported by a KoUP funding of the University of Potsdam and by the Auxological Society.

References

- Corley, J./Gow, A. J./Starr, J. M./Deary, I. J. (2010). Is body mass index in old age related to cognitive abilities? The Lothian birth cohort 1936 study. *Psychology and Aging* 25 (4), 867–875. <https://doi.org/10.1037/a0020301>.
- Durant, W. (1991). *Story of philosophy*. 2nd ed. New York, Simon & Schuster.
- Gasparatos, N./Scheffler, C./Hermanussen, M. (2023). Assessing the applicability of changepoint analysis to analyse short-term growth. *Human Biology and Public Health* 3, 1–2. <https://doi.org/10.52905/hbph2023.1.62>.

Groth, D./Scheffler, C./Hermanussen, M. (2019). Body height in stunted Indonesian children depends directly on parental education and not via a nutrition mediated pathway – Evidence from tracing association chains by St. Nicolas House Analysis. *Anthropologischer Anzeiger*. <https://doi.org/10.1127/anthranz/2019/1027>.

Groth, Detlef (2023). snha: Package for creating correlation networks using the St. Nicolas House Algorithm. Potsdam, Germany 2023. Available online at <https://github.com/mittelmark/snha> (accessed 6/21/2023).

Hake, T./Bodenberger, B./Groth, D. (2023). In Python available: St. Nicolas House Algorithm (SNHA) with bootstrap support for improved performance in dense networks. *Human Biology and Public Health* 3, 1–2. <https://doi.org/10.52905/hbph2023.1.63>.

Hermanussen, M./Geiger-Benoit, K./Burmeister, J./Sippell, W. G. (1988). Periodical changes of short term growth velocity ('mini growth spurts') in human growth. *Annals of human biology* 15 (2), 103–109. <https://doi.org/10.1080/03014468800009521>.

Malambo, C./Musalek, M. (2023). No association between anthropometry and IQ in Czech preschool children. *Human Biology and Public Health* 3, 1–2. <https://doi.org/10.52905/hbph2023.1.65>.

R Core Team (2022). R: A language and environment for statistical computing. Vienna, Austria. Available online at <https://www.R-project.org/> (accessed 6/21/2023).

Rösler, A./Scheffler, C./Hermanussen, M. (2023). No evidence of growth impairment after forced migration in Polish school children after World War II. *Human Biology and Public Health* 3. <https://doi.org/10.52905/hbph2023.1.68>.