The Journal of Open Source Software

BGmisc: An R Package for Extended Behavior Genetics Analysis

S. Mason Garrison[1], **Michael D. Hunter**[2], **Xuanyu Lyu**[1], **Jonathan D. Trattner**[1], and **S. Alexandra Burt**[3]

1 Wake Forest University, North Carolina, USA **2** Pennsylvania State University, Pennsylvania, USA **3** Michigan State University, Michigan, USA

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Behavior genetics focuses on understanding genetic and environmental influences on individual differences, traditionally through twin studies. However, with the expansion of research to more complex data structures like extended family data, there arises a need for specialized software tools. The `BGmisc` package addresses this gap by offering a comprehensive suite of functions for structuring and modeling such data.

Statement of need

As behavior genetics delves into more complex data structures like pedigrees, the limitations of current tools become evident. The `BGmisc` R package addresses these challenges, going beyond what is available in tools like `OpenMx` and `EasyMx` that mainly focus on classical twin models.

Two widely used R packages in genetics modeling are `OpenMx` (Neale et al., 2016) and `kinship2` (J. P. Sinnwell et al., 2014; J. Sinnwell & Therneau, 2022). The `OpenMx` (Neale et al., 2016) package is a workhorse in behavior genetic research. Not only is it a general-purpose software for structural equation modeling that is popular among behavior geneticists (Garrison, 2018), but also for its unique features – the `mxCheckIdentification()` function. This function checks whether a model is identified, determining if there is a unique solution to estimate the model's parameters based on the observed data. In addition, `EasyMx` (Hunter, 2023) is a more user-friendly package that streamlines the process of building and estimating structural equation models. It seamlessly integrates with `OpenMx`'s infrastructure. Its functionalities range from foundational matrix builders like `emxCholeskyVariance` and `emxGeneticFactorVariance` to more specialized functions like `emxTwinModel` designed for classical twin models. Despite their strengths, `EasyMx` and `OpenMx` have limitations when handling extended family data. Notably, they lack functions for handling modern molecular designs (Kirkpatrick et al., 2021), modeling genetic complex relationships, inferring relatedness, or simulating pedigrees.

Although not a staple in behavior genetics, the `kinship2` (J. P. Sinnwell et al., 2014) package provides core features to the broader statistical genetics scientific community, such as plotting pedigrees and computing genetic relatedness matrices. It uses the Lange algorithm (Lange, 2002) to compute relatedness coefficients. This recursive algorithm is discussed in great detail elsewhere, laying out several boundary conditions and recurrence rules. The `BGmisc` package extends the capabilities of `kinship2` by introducing an alternative algorithm to calculate the relatedness coefficient based on network models. By applying classic path-tracing rules to the entire network, this new method is computationally more efficient by eliminating the need for a multi-step recursive approach.

Features

The BGmisc package offers various features tailored for extended behavior genetics analysis. These features are grouped under two main categories, mirroring the structure presented in our vignettes.

Modeling and Relatedness:

- Model Identification: BGmisc evaluates whether a variance components model is identified and fits the model's estimated variance components to observed covariance data. The technical aspects related to model identification have been described in Hunter et al. (2021).
- Relatedness Coefficient Calculation: Using path tracing rules first described in (Wright, 1922) and formalized in (McArdle & McDonald, 1984), BGmisc calculates the (sparse) relatedness coefficients between all pairs of individuals in extended pedigrees based purely on mother and father identifiers.
- Relatedness Inference: BGmisc infers the relatedness between two groups based on their observed total correlation, given additive genetic and shared environmental parameters.

Pedigree Analysis and Simulation:

- Pedigree Conversion: BGmisc converts pedigrees into various relatedness matrices, including additive genetics, mitochondrial, common nuclear, and extended environmental relatedness matrices.
- Pedigree Simulation: BGmisc simulates pedigrees based on parameters including the number of children per mate, generations, sex ratio of newborns, and mating rate.

Collectively, these tools provide a valuable resource for behavior geneticists and others who work with extended family data. They were developed as part of a grant and have been used in several ongoing projects (Burt, 2023; Garrison et al., 2023; Hunter et al., 2023; Lyu et al., 2023), and theses (Lyu, 2023).

Availability

The BGmisc package is open-source and available on both GitHub at https://github.com/R-Computing-Lab/BGmisc and the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/package=BGmisc. It is licensed under the GNU General Public License.

Acknowledgements

The current research is supported by the National Institute on Aging (NIA), RF1-AG073189. We want to acknowledge assistance from Carlos Santos.

References

Burt, S. A. (2023). Mom genes: Leveraging maternal lineage to estimate the contributions of mitochondrial DNA. *Behavior Genetics*.

Garrison, S. M. (2018). Popular Structural Equation Modeling Programs for Behavior Genetics. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 972–977. https://doi.org/10.1080/10705511.2018.1493385

Garrison, S. M., Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., & Burt, S. A. (2023). Analyzing extended cousin similarity to unravel the mystery of mtDNA and longevity. *Behavior Genetics*.

Hunter, M. D. (2023). *EasyMx: Easy model-builder functions for 'OpenMx'*. https://CRAN.R-project.org/package=EasyMx

Hunter, M. D., Garrison, S. M., Burt, S. A., & Rodgers, J. L. (2021). The Analytic Identification of Variance Component Models Common to Behavior Genetics. *Behavior Genetics*, *51*(4), 425–437. https://doi.org/10.1007/s10519-021-10055-x

Hunter, M. D., Lyu, X., Garrison, S. M., Rodgers, J. L., Smith, K., Coon, H., & Burt, S. A. (2023). Modeling mtDNA effects from extended pedigrees in the utah population database. *Behavior Genetics*.

Kirkpatrick, R. M., Pritikin, J. N., Hunter, M. D., & Neale, M. C. (2021). Combining Structural-Equation Modeling with Genomic-Relatedness-Matrix Restricted Maximum Likelihood in OpenMx. *Behavior Genetics*, *51*(3), 331–342. https://doi.org/10.1007/s10519-020-10037-5

Lange, K. (2002). Genetic Identity Coefficients. In K. Lange (Ed.), *Mathematical and Statistical Methods for Genetic Analysis* (pp. 81–96). Springer. https://doi.org/10.1007/978-0-387-21750-5_5

Lyu, X. (2023). *Statistical power analysis on mtDNA effects estimation* [Master's thesis]. Wake Forest University.

Lyu, X., Hunter, M. D., Rodgers, J. L., Smith, K. R., Coon, H., Burt, S. A., & Garrison, S. M. (2023). Statistical power analysis on mtDNA effects estimation. *Behavior Genetics*.

McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the reticular action model for moment structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 234–251. https://doi.org/10.1111/j.2044-8317.1984.tb00802.x

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/f8rfrg

Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 r package for pedigree data. *Human Heredity*, *78*, 91–93. https://doi.org/10.1159/000363105

Sinnwell, J., & Therneau, T. (2022). *kinship2: Pedigree functions*. https://CRAN.R-project.org/package=kinship2

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, *56*(645), 330–338. https://doi.org/10.1086/279872