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**[1,3,4], **Jonathan D. Trattner**[1], and **S. Alexandra Burt**[5]

1 Department of Psychology, Wake Forest University, North Carolina, USA **2** Department of Human Development and Family Studies, Pennsylvania State University, Pennsylvania, USA **3** Institute for Behavioral Genetics, University of Colorado at Boulder, Boulder, USA **4** Department of Psychology & Neuroscience, University of Colorado at Boulder, Boulder, USA **5** Department of Psychology, Michigan State University, Michigan, USA ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

The field of behavior genetics focuses on illuminating genetic and environmental influences on individual differences. Traditionally, twin studies have been at the forefront of this discipline. However, research has moved beyond the classical twin design to embrace more complex family structures such as children of twins (CoT) (D'Onofrio et al., 2003), mother-daughter-aunt-niece (MDAN) (Rodgers, Bard, Johnson, D'Onofrio, & Miller, 2008), and other extended family designs. These expansions allow for a deeper, more nuanced exploration of genetic and environmental influences, but it also introduces challenges, particularly in data structuring and modeling. In particular, the data structures inherent in these more complicated family designs are orders of magnitude larger than traditional designs. In the classical twin study, for example, a family will consist of a single pair of twins (i.e., two people), whereas in the MDAN design, a family consists of two mother-daughter pairs (i.e., four people). This problem quickly becomes intractable when applied to very extended family pedigrees, where a single family can be of any size. The `BGmisc` package addresses this gap by offering a comprehensive suite of functions for structuring and modeling extended family pedigree data.

Statement of need

As behavior genetics delves into more complex data structures like extended pedigrees, the limitations of current tools become evident. Understandably, packages like `OpenMx` (Neale et al., 2016), `EasyMx` (Hunter, 2023), and `kinship2` (J. P. Sinnwell, Therneau, & Schaid, 2014; J. Sinnwell & Therneau, 2022) were built for smaller families and classical designs. In contrast, the `BGmisc` R package was specifically developed to structure and model extended family pedigree data.

Two widely-used R packages in genetic 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 general-purpose software for structural equation modeling that is popular among behavior geneticists (Garrison, 2018) for its unique features, like 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, Pritikin, Hunter, & Neale, 2021), modeling complex genetic relationships, inferring relatedness, and 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 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 by Hunter, Garrison, Burt, & Rodgers (2021).
- Relatedness Coefficient Calculation: Using path tracing rules first described by Wright (1922) and formalized by McArdle & McDonald (1984), `BGmisc` calculates the (sparse) relatedness coefficients between all pairs of individuals in extended pedigrees based solely 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.

Acknowledgments

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*.

D'Onofrio, B. M., Turkheimer, E. N., Eaves, L. J., Corey, L. A., Berg, K., Solaas, M. H., & Emery, R. E. (2003). The role of the children of twins design in elucidating causal relations between parent characteristics and child outcomes. *Journal of Child Psychology and Psychiatry*, *44*(8), 1130–1144. doi:ff3dbm

Garrison, S. M. (2018). Popular Structural Equation Modeling Programs for Behavior Genetics. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 972–977. doi: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'*. Retrieved from 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. doi: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. doi:10.1007/s10519-020-10037-5

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

Lyu, X. (2023, May). *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. doi: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., et al. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. doi:f8rfrg

Rodgers, J. L., Bard, D. E., Johnson, A., D'Onofrio, B., & Miller, W. B. (2008). The Cross-Generational Mother–Daughter–Aunt–Niece Design: Establishing Validity of the MDAN Design with NLSY Fertility Variables. *Behavior Genetics*, *38*(6), 567–578. doi:10.1007/s10519-008-9225-0

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

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

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