Thom Benjamin Volker

Utrecht, The Netherlands

thomvolker.github.io | t.b.volker@uu.nl | GitHub | Twitter | LinkedIn

EDUCATION

MSc Methodology and Statistics **2019 - 2022**
Utrecht University, Netherlands *Cum laude (GPA 8.9/10)*

- Thesis - Combining support for hypotheses over heterogeneous studies with Bayesian Evidence Synthesis: A simulation study (supervised by Prof. Dr. Irene Klugkist).

MSc Sociology and Social Research **2019 - 2022**
Utrecht University, Netherlands *Cum laude (GPA 8.7/10)*

- Thesis - The future is made today: Concerns for reputation foster trust and cooperation (supervised by Prof. Dr. Ir. Vincent Buskens & Prof. Dr. Werner Raub).

BA Liberal Arts & Sciences **2016 - 2019**
Utrecht University, Netherlands

- Major in Pedagogical Sciences; minor in Sociology and Social Research

EXPERIENCE

PhD candidate Synthetic data **2020 - Current**
Utrecht University

- Supervisor: Dr. Erik-Jan van Kesteren
- I investigate how to optimize the creation of synthetic data in terms of the risk-utility trade-off.

Internship MICE Group **2020 - 2022**
Utrecht University

- With Dr. Gerko Vink I established a framework on how to generate synthetic data with the `R`-package `mice` to preserve privacy and confidentiality.

Research Assistant **2019 - 2022**
Utrecht University

- For Dr. Peter Lugtig, I worked on research and teaching related topics, (e.g., creating content for the research master level course 'Survey Data Analysis', by developing a Shiny App, creating data visualizations and creating the webpage https://www.peterlugtig.com.
- For Dr. Rebecca Kuiper, I created data visualizations and assisted PhD students by revising text.

Teaching assistant **2018 - 2022**
Utrecht University

- Thesis co-supervision:
  - Multiply imputed synthetic datasets - assessing validity of multiply imputed synthetic datasets with the `R`-package `mice` (Master's thesis, with Dr. Gerko Vink).
  - Comparing multiple methods of research synthesis (Bachelor's thesis, with Prof. Dr. Irene Klugkist).
- Post-graduate level courses:
  - Multiple Imputation in Practice (MIMP; Utrecht Summer School 2019-2021) - supervision of practicals and providing `R` assistance.
  - Statistical Programming with R (Utrecht Summer School 2021) - supervision of practicals.
  - Advanced Survey Design - supervision of practicals.
- Master's level courses:
  - Network Analysis (2020-2021) - developing practical assignments and supervising the practicals.
  - Methodological and Statistical Aspects of Social Science Research - teaching working groups of about 20 persons on topics ranging from multiple regression to moderation and mediation analysis.
- Bachelor's level courses:

- Theory Construction and Statistical Modeling - teaching practicals on structural equation modeling using `lavaan`.
- Various undergraduate courses on standard statistical methods.

ASReview **2021**

Utrecht University

- Researcher at ASReview - an open-source project that helps researchers to screen (tens of) thousands of papers automatically for inclusion in systematic reviews, meta-analyses, medical guidelines, or overviews.

Other

Debuut **2019 - 2020**

Utrecht University

- Buddy programme at Utrecht University to match potential first generation university student with actual students to get a flavour of what studying at a university entails.

Statistical software proficiency

- `R`
- `MPlus`
- `SPSS`
- `HLM`
- `JAGS`
- Learning `python`
- Learning `C++`

Publications

Volker, T. B., & Vink, G. (2021). Anonymiced shareable data: Using mice to create and analyze multiply imputed synthetic datasets. *Psych*, *3*(4), 703–716. doi: 10.3390/psych3040045

Software Vignettes

Volker, T. B. & Vink, G. (2022). `futuremice`: The future starts today. `https://www.gerkovink.com/futuremice/Vignette_futuremice.html`

Software development

- `mice` [ctb] - Developed functionality to pool inferences from synthetic data and implemented `futuremice` to speed up imputation by using user-friendly parallel processing.
- `ggmice` [ctb] - Adjusted functionality to allow for visualization of synthetic data.