**Software preferences for modeling preferences: where to StaRt?**  
*27801/PSPE*   
  
**Natalie Riva Smith1, Norah L Crossnohere2, Tom Chen3, Davene R Wright3**  
1Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, USA  
2Division of General Internal Medicine, Department of Internal Medicine, The Ohio State University College of Medicine, Columbus, OH, USA  
3Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA, USA

**Purpose:** We compared software options for analyzing stated-preference data to help researchers make informed decisions about which is the most appropriate for their project.

**Methods:** We focus on two analysis platforms common in the medical decision-making literature, Stata and R. Within R, we compare three packages with varying functionalities: apollo, logitr, and mlogit. There is minimal guidance on how to compare statistical software. Using our team’s experiences in health services, preferences, econometrics, and biostatistics, we selected three key domains for comparison: available functionality, computational considerations, and accessibility. To demonstrate the use of each software option, we estimated multinomial, mixed logit, and latent class logit models using data from a discrete choice experiment examining health insurance choice among the US population aged 18-64.

**Results:** Available functionality was similar across most software options. All can estimate multinomial and mixed logit models, while only Stata and R/apollo have clear procedures for estimating latent class logit models. Computational considerations (e.g., input data formats, approximation of marginal likelihoods, computation time) and accessibility (e.g., cost, accessing help) differed across software options (Table). Analytic results were generally consistent, though estimates in mixed logit models varied slightly due to differences in the Monte Carlo approximation of the marginal likelihood. Annotated code to conduct analyses in each software option will be provided in a GitHub repository.

**Conclusion:** Researchers should consider the trade-offs between software options (e.g., computation time vs. coding complexity) and select the software that best aligns with their functionality, computation, and accessibility needs, particularly when accounting for their broader analytic workflow and familiarity with different statistical software.

**Keywords:**patient preference methods, statistical software

**Table**: Comparisons of software options for stated preference data analysis

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Domain | Stata | R/apollo | R/logitr | R/mlogit |
| **Computational considerations** | | | | |
| Data format required | Tabular data; person-by-task-by-alternative | Tabular data; person-by-task | Tabular data; Person-by-task-by-alternative | Must coerce to dfidx object |
| Approximation of marginal likelihood for MXL | Hammersley; Halton; pseudo-random | Halton; MLHS; pseudo-Monte Carlo; Sobol | Halton; Sobol | Halton; pseudo-random |
| Computation time for MXL with 100 draws for likelihood approximation | 4.5 minutes | 1.3 minutes | 32.6 seconds | 53.5 seconds |
| **Accessibility** | | | | |
| Cost | University or individual license | | Free / open source | |
| Help | Stata forums | Online, forum | Online, Github issues | Online |
| Citation | StataCorp, 2023 | Hess and Palma, 2019 | Helveston, 2023 | Croissant, 2020 |

Notes: MXL=mixed logit; MLHS=Modified Latin Hypercube Sampling.