Software preferences for modeling preferences: where to StaRt?

Natalie Riva Smith (1), Norah L Crossnohere (2), Tom Chen (3), and Davene R Wright (3)
(1) Department of Social and Behavioral Sciences, Harvard TH Chan SPH; (2) Division of General Internal Medicine, Ohio State University; (3) Department of Population Medicine, Harvard Pilgrim Health Care Institute

Why?
Many analytic options are available for researchers when analyzing stated preference data. We compared software options for analyzing stated-preference data to help researchers make informed decisions about which is the most appropriate for their project.

How?
- Compared functionality, computational considerations, and accessiblity.
- 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.

To select a software that best aligns with your needs, consider your:

Analytic workflow
Functionality needs
Computational needs
Software accessibility
Familiarity with software
Personal preference

So... what to choose?

	Stata	R / apollo	R / logitr	R / mlogit
Data format	Tabular Person-by-task-by-alternative	Tabular Person-by-task	Tabular Person-by-task-by-alternative	Coerce to `dfidx` object
Support for survey data	`svy` prefix	Person weights and robust standard errors	Person weights and robust standard errors	Person weights and robust standard errors
Modeling capabilities*	CL, MXL, NL, OL (cm suite) LCL via additional function	CL, MXL, NL, OL, LCL	CL, MXL	CL, MXL, NL, OL
Approximation of marginal likelihood for MXL	Hammersley; Halton; psuedo-random	Halton; MLHS; pseudo-Monte Carlo; Sobol	Halton; Sobol	Halton; pseudo-random
Computation time in MXL**	4.5 minutes	1.3 minutes	32.6 seconds	53.5 seconds
Key tradeoffs to consider	Established software familiar to many MDM researchers, but with a financial and computational power price	Extreme flexiblity in modeling and robust online community, but comes with a higher learning curve	Fast, user-friendly, and integrates with `tidyverse`, but more limited in modeling capabilities	Alternative R package that may be more easily integrated into workflows depending on processes

* We asked whether the softwares had the capability to run conditional logit (CL), mixed logit (MXL), nested logit (NL), ordered logit/probit (OL), and latent class logit models (LCL); ** MXL with 100 draws for likelihood approximation

HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH
THE OHIO STATE UNIVERSITY COLLEGE OF MEDICINE
Harvard Pilgrim Health Care Institute
ORCHID Optimizing Choice of Health Insurance for Diabetes

Scan to see our code for this poster and connect!

Acknowledgements
Data used in these analyses were collected as part of a Pyle Fellowship to NRS by the Harvard Pilgrim Health Care Institute. NRS is supported by K99CA277135.