Pre-course Reading

Workshop on Analyzing Mixtures in Environmental Health Studies

To get the most out of the Environmental Mixtures Workshop, please read the required paper below **prior to August 19th**. We've also included some optional reading if you would like to further prepare.

Please note that readings were assigned letters as follows: **T – Theoretical; M – Methodological; A – Applied.**

Details below:

Required reading

The workshop will be using the dataset described in Mitro et al. 2016:

1. Mitro SD, Birnbaum LS, Needham BL, Zota AR. Cross-sectional Associations between Exposure to Persistent Organic Pollutants and Leukocyte Telomere Length among U.S. Adults in NHANES, 2001–2002. Environmental Health Perspectives. 2016;124(5):651-658. DOI:10.1289/ehp.1510187.

Optional reading: Primary

This paper gives a good introduction to and overview of statistical methods.

1. Gibson EA, Goldsmith J, Kioumourtzoglou MA. Complex Mixtures, Complex Analyses: An Emphasis on Interpretable Results. Current environmental health reports. 2019;6(2):53-61. **[Review Paper]**

2. Carrico C, Gennings C, Wheeler DC, Factor-Litvak P. Characterization of weighted quantile sum regression for highly correlated data in a risk analysis setting. Journal of Agricultural, Biological, and Environmental Statistics. 2015;20(1):100-20. **[M]**

3. Chapters 6.2 (Lasso) and 10 (PCA and clustering):
James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. *New York: Springer*. 2013. **[M]**

4. *BKMR modeling framework:* Bobb JF, Valeri L, Claus Henn B, Christiani DC, Wright RO, Mazumdar M, Godleski JJ, Coull BA. Bayesian kernel machine regression for estimating the health effects of multi-pollutant mixtures. *Biostatistics (Oxford, England)*. 2015;16(3):493-508. DOI:10.1093/biostatistics/kxu058. **[M]**

5. *Software implementation of BKMR methods in R:* Bobb JF, Claus Henn B, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environmental Health*. 2018;17(1):67. DOI: 10.1186/s12940-018-0413-y. **[A]**

Page 2

Optional reading: Secondary

1. Curtin P, Kellogg J, Cech N, Gennings C. A random subset implementation of weighted quantile sum (WQS_{rs}) regression for analysis of high-dimensional mixtures. Communications in Statistics - Simulation and Computation. 2019;1-16. DOI: 10.1080/03610918.2019.1577971. **[A]**

2. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001;96(456):1348-1360. **[T]**

3. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological).* 1996;58(1):267-288. **[T]**

4. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 2006;68(1):49-67. **[T]**

5. Zhang C. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*. 2010;38(2):894-942. **[T]**

6. Liu SH, Bobb JF, Lee KH, Gennings C, Claus Henn B, Bellinger D, Austin C, Schnaas L, Tellez-Rojo MM, Hu H, Wright RO, Arora M, Coull BA. Lagged kernel machine regression for identifying time windows of susceptibility to exposures of complex mixtures. Biostatistics. 2018;19(3):325-341. **[M]**

7. Liu SH, Bobb JF, Claus Henn B, Gennings C, Schnaas L, Tellez-Rojo MM, Bellinger D, Arora M, Wright RO, Coull BA. Bayesian varying coefficient kernel machine regression to assess neurodevelopmental trajectories associated with exposure to complex mixtures. Stat Med. 2018;37(30):4680-4694. **[M]**

8. Liu JZ, Lee J, Lin PI, Valeri L, Christiani DC, Bellinger D, Wright RO, Mazumdar MM, Coull BA. A Cross-validated Ensemble Approach to Robust Hypothesis Testing of Continuous Nonlinear Interactions: Application to Nutrition-Environment Studies. arXiv:1904.10918. 2019. **[M]**

9. Antonelli J, Mazumdar MM, Bellinger D, Christiani DC, Wright RO, Coull BA. Estimating the health effects of environmental mixtures using Bayesian semiparametric regression and sparsity inducing priors. arXiv:1711.11239. 2017. **[M]**