

Case Study 1-Group 1

Melody Jiang, Irene Ji, Keru Wu

Department of Statistical Science, Duke University

01/21/2019

Introduction

- ▶ Data: A study by Longnecker et al. (2001), comprised of 2380 observations of pregnant women.
- ▶ Goal: Assess how DDE and PCBs relate to risk of premature delivery.

Data Preprocessing

- ▶ Preterm birth: Gestational Age ≤ 36 .
- ▶ Standardize continuous variables.
- ▶ Missing data: Multivariate Imputations by Chained Equations (MICE package in R).
- ▶ Limit of Detection (LOD): Exists in some PCBs. All LODs are negligible compared to data scale (e.g. 0.01 compared to 0.3)

Collinearity & Dimensionality Reduction

- ▶ There are 11 types of PCBs, some of which have high correlation and may be redundant.
- ▶ Possible approaches: Simple sum, PCA, Factor Analysis .
- ▶ Alternative approach: model with regularization (e.g. LASSO).
- ▶ Check Multicollinearity: Variation Inflation Factor (VIF).

Model

- ▶ Generalized Additive Model (GAM)

$$g(Y_i) = \beta_0 + \sum_{j=1}^m f_j(x_{ij}) + \sum_{k=1}^l \beta_k z_{ik}$$

- ▶ Choice of g : probit or logit.
- ▶ $x_{.j}$ s include DDE, PCBs.
- ▶ $z_{.k}$ s include categorical variables and confounding variables.

Model

- ▶ Frequentist model overestimates uncertainty.
- ▶ Bayesian Generalized Additive Model

$$g(Y_i) = \beta_0 + \sum_{j=1}^m f_j(x_{ij}) + \sum_{k=1}^l \beta_k z_{ik}$$

- ▶ Adds priors on the common regression coefficients, priors on the standard deviations of the smooth terms, and a prior on the decomposition of the covariance matrices of any group-specific parameters

Discussion

- ▶ Monotone effect: Bayesian Isotonic Regression (Neelon, B. and Dunson, D.B. 2004)
- ▶ Including Interactions: Bayesian Factor Analysis (Ferrari, F. and Dunson, D.B. 2019)