Assessing Effects of Exposures to DDE and PCBs on Premature Delivery via Ordinal Logistic Regression

Raphael Morsomme Rihui Ou Alessandro Zito

Case Study 1- Stat 723

January 19, 2020

Overview

Introduction

- **Framework**:
  Dichlorodiphenyldichloroethylene (DDE) and *Polychlorinated Biphenyls* (PCBs) are chemicals that persist in the envirnoment and get stored in fatty depositis in the human tissues.
  $\Longrightarrow$ Potential adverse effect on health
- **Question**:
  Is exposure to DDE and PBCs associated with a higher chance of premature delivery in pregnant women?

Pregnancy timeline

- **Dangerous preterm**: delivery at 34 weeks or before (when main organs are underdeveloped)
- **Preterm**: delivery beween 35 and 37 week
- **At term**: delivery after 37 weeks

Data

Data contained gestational age (in weeks) of the mother, the DDE and PCBs concentration, socio-economic info and scores (race, occupation, education and income), and amount of triglycerides and cholesterol. Total sample size (after preprocessing) = 2336
We contruct the following variables:

- Total level of lipids[1]

$$lipid_i = 2.27 * cholesterol_i + triglycerides_i + 0.623$$

- Gestational age group

$$gestgroup_i = \begin{cases} 0 & if \text{ Dangerous preterm} \\ 1 & if \text{ Preterm} \\ 2 & if \text{ At term} \end{cases}$$

- Average (standardized) PCB

$$PCB_i = \frac{1}{11} \sum_{j=1}^{11} \frac{PCB_{ij} - mean_i(PCB_{ij})}{sd_i(PCB_{ij})}$$

[1] Using Phillips et al.(1989) and Bernert et al.(2007)

Model (I) - Ordinal Logistic Regression

We run the following ordinal logistic regression model:

$$\log \frac{P(gestgroup_i \leq j)}{P(gestgroup_i > j)} = \beta_{0j} - \eta_1 \frac{DDE_i}{lipid_i} - \eta_2 \frac{PCB_i}{lipid_i} - \boldsymbol{\xi}^T \mathbf{z}_i + \varepsilon_i$$

where

- $j = 0, 1, 2$ is the outcome level
- DDE_i and PCB_i are the amount of DDE and PCB
- $lipid_i$ measures the lipid deposit
- $\mathbf{z}_i$ is a set of covariates.

After an AIC backward variable selection procedure, we determine that $\mathbf{z}_i = (center_i, score\_education_i)$
Model assumptions are checked in the appendix.

Model (II) - Bayesian Ordinal Logistic Regression

Results

Conclusions

Appendix

Preprocessing:

- Drop obs. with gestational age > 45 (the world record)
- Standardize and average the differnt PCBs (to avoid their correlation)
- Mean impute of occupation, education and income scores $\Longrightarrow$ Total obs. $= 2336$

Model Checking

We can check the assumption of the (frequentist) ordinal logistic model by looking at the Surrogate residuals. If the model assumptions are correct, then the surrogate residuals R_S will have three properties:

- $E(R_S|X) = 0$
- $Var(R_S|X) = c$, the conditional variance of R_S is constant
- The emiprical distribution of R_S resembles an explicit distribution that is related to the link function $G^{-1}(\cdot)$. Specifically, $R_S \sim G(c + \int u dG(u))$.