Quantile regression with applications to biomonitoring of human exposure to environmental chemicals (Parametric Approach)

By

Wu Yue

Supervisor:

Prof. Anthony Kuk Yung Cheung

ST4199   Honours Project in Statistics

Department of Statistics and Applied Probability

National University of Singapore

2014/2015

Abstract

Biomonitoring is the measurement of the body burden of toxic chemical compounds, elements, or their metabolites. Chemicals or their metabolites were measured in blood and urine samples from a random sample of participants from the National Health and Nutrition Examination Survey (NHANES) conducted by CDC's National Center for Health Statistics. NHANES is a series of surveys designed to collect data on the health and nutritional status of the U.S. population [1]. In this report, the NHANES 2003-2004 data set is used for analyzing the exposure situations of PCB153 as a function of age as well as effects of race and gender.

A parametric approach is adopted as we assume that the measurements follow a log-normal distribution, which is a common assumption in the biomonitoring literature. Furthermore, quantile regression, as opposed to regression of the mean, is used to explore the conditional distribution of a response variable given the values of one or more covariates [2]. In addition, we also consider the generalized gamma model which includes the log-normal model as a limiting case.

The whole paper consists of model selection regarding to the whole dataset, subgroup quantile regression analysis, quadratic term inclusion using specific tests, confidence band analysis for different quantiles, simulations for testing the correctness of confidence band coverage as well as the comparison between the log-normal and generalized gamma fit.