

ST740 – Midterm Exam 2

Due in class, Monday, 11/7/11

Your second midterm is a take-home exam. You will conduct a Bayesian data analysis and write a 5-6 page summary as described below. The paper is due at the beginning of class on 11/7; no late work will be accepted. Also, no communication with anyone regarding this paper is allowed. If you have any questions about the data or my expectations, please email me or stop by my office, but do not seek outside advice.

For this project, you will analyze the data in the R workspace “NCbirth.RData” on the course webpage. Once you have downloaded the data to your computer, you can load the data in R by clicking file → load workspace, and browsing for the file on your PC. The data are a random subset of births in North Carolina in 2010. Each observation is a birth. The binary response is “preterm”. The remaining variables are predictors you may include in your model. A description of the variables is given at the end of this assignment.

Your objective is to build a statistical model using Bayesian methods for preterm birth and address the following research questions: Which county has the highest rate of preterm birth? Does preterm birth rate vary by race (white or not white)? If there is a racial difference in preterm birth rate, which county has the largest disparity?

In your final report, please include the following sections:

- **Introduction:** Very briefly state the objectives of your paper and define the data and scope of your analysis.
- **Methods:** Describe the statistical model or models you use and argue why they are appropriate. Also give all priors and the methods used to select the final model.
- **Computing details:** Describe the computational algorithms you used to analyze the data, and provide some supporting evidence that the algorithms were effective.
- **Results:** Present the main results of the analysis, and discuss whether the results are sensitive to modeling assumptions or prior distributions.
- **Conclusions:** Summarize your main findings and relate them to the objectives given in the introduction. Also, discuss any limitations to your analysis and directions of future work.

Your paper should be 5-6 pages (double-spaced, including tables and figures), typed in either word or latex, and written in manuscript style with numbered sections, figures, equations, etc. You do not need to turn in computer code. Please bring 3 copies to class, one with your name to be graded by me and two without your name for peer review. All three should have the same “code” on the top so that I can return them all to you.

HAVE FUN!

VARIABLE DESCRIPTIONS

preterm: 1 = birth was pre-term, 0 = birth was full-term

county: Index of the county of residence. County names are in “county_name”, so that mother i lives in county `county_name[county]`.

age_F: Father’s age

age_M: Mother’s age

edu_F: Father’s years of education

edu_M: Mother’s years of education

race_F: 1 = father is Caucasian, 0 otherwise

race_M: 1 = mother is Caucasian, 0 otherwise

smoke: 1 = mother reported smoking during pregnancy, 0 otherwise

drink: 1 = mother reported using alcohol during pregnancy, 0 otherwise

gender: 1 if baby was male, 0 otherwise