Statistical Modeling and Methods: Final Project

Due by **11:59pm of June 12, 2020**, online through Blackboard

Policy: You are NOT allowed to share your models, analysis for specific problems, codes, or outputs. Submit all your computer codes and necessary outputs.
Grading: Project will be marked according to intellectual quality, appropriateness and correctness of your answers and analysis, justifications of your methods and models, as well as the presentation of the report (conveying the main points while being concise).

1. The following data were obtained in a dose-response experiment where a vaccine was applied, reported here as triplets (log(dose), group size exposed, number of responses where vaccine was effective):
(.71,49,16), (1.0,48,18), (1.31,48,24), (1.48,49,37), (1.61,50,39), (1.7,48,40).
Fit logistic models to the data with log(dose) as predictor, use the Aranda-Ordaz family of link functions and the embedding technique to test whether a logitic link would be adequate.

2. A new treatment (coded as 1) was compared to a standard treatment (coded as 0). The dataset CONV.DAT contains the group membership (column 1), the dose levels in original scale (column 2), the number of observations at each dose level (column 3) and the corresponding number of responses (column 4). Let $\mathrm{ED}\tilde{\alpha}$ be the effective dose level of $\tilde{\alpha}$ percent, where $\tilde{\alpha} = 100\alpha$, and the quotient $q(\alpha)$ is the relative strength of the new treatment to the standard one, i.e., $q(\alpha) = \mathrm{ED}_1\tilde{\alpha}/\mathrm{ED}_0\tilde{\alpha}$. If a probit link function is used, test the log-parallerity between effective levels of these two drugs, and provide a precision measure on the relative strength of the two drugs with 95% confidence.

3. (a) For the Byssinosis data LUNG.dat, start with a binomial model using the logistic link including three relevant covariates (dust level, smoking, length of employment). Apply nonparametric regression techniques to estimate the variance function $v(\mu)$ appropriately. Is this non-parametrically estimated variance function compatible with the binomial model assumption? Explain how you estimate the variance function with suitably selected tuning parameters.

   (b) Construct the Quasi-likelihood estimates for the regression coefficients, using the non-parametrically estimated variance function, and compare these estimates to those obtained in the original logistic model. Comment on your findings.

Instruction for Data Analysis: *You are expected to provide a concise data analysis report, including the main text (not exceeding 3 pages) and an appendix (computer outputs, plots and codes etc.). You need to include a brief introduction, methods, results and conclusion sections in the report. Exploratory data inspection and model diagnostics are suggested. Describe the methods that lead to the results with appropriate justifications. The interpretation of your results should be oriented towards the audience without much statistical background. Material in the appendix need to be referenced appropriately in the text when necessary. The write-up of the report will also be taken into account for assigning credit.*

4. Historical records are available on the fertility of a cohort of French-Canadian women. A subset of these data with selected variables is in FRCAN.dat. The data is organized as follows:

   col 1: marriage id

   col 2: French-Canadian immigrant=1; Non-immigrant=0;

   col 3: husband id;

   col 4: date of death of current husband; -99 if missing;

   col 5: woman id;

   col 6: number of marriages;

   col 7: date of birth of woman (yyyymmdd);

   col 8: 1 if missing exact birth date; 0 otherwise;

   col 9: date of death of woman (yyyymmdd);

   col 10: 1 if missing exact death date; 0 otherwise;

   col 11: number of children.

   Your task is to model the number of children and relate it to the predictors which also need to be identified and more precisely specified. Possible relevant predictors are lifespan, number of marriages and immigrant status, among others.