Stat 545 Part II Homework # 3

Fall 2015 Rice University

Instructor:	Liang Li <LLi15@mdanderson.org>
Teaching Assistant:	Yinsen Miao <ym11@rice.edu>
Date of Assignment:	Nov 23, 2015
Due Date:	December 2, 2015

Instructions: Please prepare your answers in Microsoft Word (with Equation Editor) or Latex. Please include your SAS (preferred) or R code, copy and paste the raw output from the software, and interpret the results.

Problem # 1 (*10 points*). Please use conditional logistic regression to analyze the data in Table 11.1 (page 414) of the textbook. The model should be specified as the following: $\text{logit}\left[P(Y_{ij}=1)\right] = \alpha_i + \beta \times j$, where $i = 1, 2, ..., 433$ is the index of voters, $j = 0$ or 1 is the index for year 2004 or 2008 respectively, and $Y_{ij} = 1$ if the vote is for Democrats, and 0 for Republican.

1. Please fit this conditional logistic model using either SAS or R. Report the code and output. Interpret the result on β. Is there any evidence that the voter preference changed from 2004 to 2008? Did voter preference for Democrats increase or decrease?
2. What does α_i represents?
3. Please calculate the maximum likelihood estimate of the population averaged effect and the maximum conditional likelihood estimate of β in the subject-specific model above (Hint: read §11.2). Are they the same? If not, why? (Hint: read §13.2.3)
4. Use SAS or R to perform a McNemar's test, and compare the result with conditional logistic regression.
5. Suppose the data have been changed in such a way that there are 100 voters who voted for Democrats in both 2004 and 2008, and 263 voters who voted for Republicans in both 2004 and 2008. Please rerun the conditional logistic model and McNemar's test. Do the results change? Please explain why.

Problem # 2 (*10 points*). For the attitudes about abortion data of Table 11.13 in Section 11.7.4, consider the model

$$logit[P(Y_t = 1)] = \alpha + \beta_1 I(t = 1) + \beta_2 I(t = 2) + \beta_3 X$$

where $t = 1, 2, 3$ indicating the three questions (items) in the footnote of Table 11.13. $I(t = 1)$ and $I(t = 2)$ are the dummy variables indicating the first and second item. X is gender, with $X = 1$ indicating females and $X = 0$ indicating males. The data are provided with this homework assignment.

1. Fit a logistic regression model by treating the three observations for a subject as independent.

2. Fit a GEE logistic regression model by using (1) exchangeable and (2) unstructured correlation structure. Report the results, including regression coefficients, their model-based and empirical standard errors, p-values, and the estimated working correlation matrix