

## ST740 – Assignment 5 – Due 11/2

For this assignment you will analyze the North Carolina election data available at

<http://www4.stat.ncsu.edu/~reich/st740/Votes08.csv>  
<http://www4.stat.ncsu.edu/~reich/st740/NCADJ.csv>.

The response for each county is the total number of votes and number of votes for Obama in the 2008 Presidential election. Your objective is to build a predictive model for these data using county-level covariates (percent in poverty, percent uninsured, percent black, percent Hispanic, median age, and percent over 65), and identify important predictors of support for Obama. You should use some variant (e.g., an arcsin transformation) of the conditionally autoregressive model

$$y_i \sim \text{Binomial}(n_i, p_i)$$

where  $y_i$  is the number of votes for Obama from county  $i$ ,  $n_i$  is the total number of votes in county  $i$ ,

$$\text{logit}(p_i) = \theta_i + \mathbf{x}_i\boldsymbol{\beta},$$

$\theta_i$  is the spatial effect,  $\mathbf{x}_i$  are covariates and  $\boldsymbol{\beta}$  are the corresponding coefficients. The spatial effects follow the CAR model with full conditional

$$\theta_i | \theta_j, j \neq i \sim N(\bar{\theta}_i, \sigma^2/m_i),$$

where  $\bar{\theta}_i$  is the mean of  $\theta$  at county  $i$ 's  $m_i$  neighbors.

Write a 2-4 page summary of your data analysis in word or latex. A latex example can be found here, <http://www4.stat.ncsu.edu/~reich/st810A/>. Your summary should include:

- Brief description and the data and objectives
- Description of the model(s) you fit to the data
- Description of MCMC algorithm and MCMC convergence diagnostics
- Model selection/adequacy/diagnostic results
- Prior sensitivity analysis
- Summary of main results

Please write your report as a short paper, using complete paragraphs, section headings, labeled figures, etc. A large portion of your grade on this assignment will be attributed to spelling, grammar, and clarity.