MLE HW #2
Deadline: Friday, February 5th, 10 AM

Use the provided dataset "County Vote for McCain" for this assignment (Dataset uploaded on Sakai > R Notes from Lecture). The variables in this dataset are as follows:

countyid: a unique indicator variable for U.S. counties
pmccain: county-level proportion two-party vote for McCain in 2008
medinc: county median income
pcollege: county proportion with college degree
punemp: county proportion unemployed
passist: county proportion on public assistance
pblack: county proportion black identifiers
phisp: county proportion Hispanic identifiers
ginicnty: county Gini coefficient (measure of household income inequality in county)
logodds of McCain proportion: ln(pmccain/(1 – pmccain))

1. Estimate the following model via maximum likelihood using bbmle in R and interpret your output.

$$logodds_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \beta_0 + \beta_1 pcollege_i + \beta_2 medinc_i$$

Specifically, do the following:

a. Write an R function for the log-likelihood that can be called from mle2 and optimized to estimate the model above.
b. Use mle2 to optimize the function given the provided data.
c. Generate marginal effects on McCain's share of the two-party vote (*not* the logodds) for both predictors. That is, calculate the difference in McCain's predicted proportion comparing counties at the 95th percentile of each predictor to those at the 5th percentile, holding the other variable at its median value. [Note: you do not need to generate confidence intervals for the marginal effects for this problem].
d. Interpret each of these effects in substantive terms: what do the results say about the predictors of McCain support and their influence relative to one another? Describe the results in an intuitive way with respect to the scales

of the predictors, such that your reader can get a sense of how these variables relate to the DV.

2. Estimate the following model via maximum likelihood using bbmle in R and interpret your output for the variance equation.

$$pmccain_i \quad N(\mu_i, \sigma^2{}_i)$$

$$\mu_i = \beta_0 + \beta_1 pcollege_i + \beta_2 medinc_i$$

$$\sigma^2{}_i = \gamma_0 + \gamma_1 ginicnty_i$$

   a. Write an R function for the log-likelihood that can be called from mle2 and optimized.
   b. Use mle2 to optimize the function given the provided data.
   c. Calculate the marginal effect of a 5-95% change in ginicnty on the standard deviation of the error term for the model.
   d. Describe and interpret this marginal effect. Does inequality have a statistically significant effect on the model errors? How do changes in Gini relate to changes in the SD? What is the substantive significance of this effect (if any)?

3. Estimate the following model via OLS (use 'lm') and use a simulation-based approach (use the 'sim' function in R) to generate point estimates and 95% confidence intervals for all predictors. Plot the estimates and their associated confidence intervals in a pretty graph, and interpret each effect in substantive terms [Note: for plotting your results, I (Chris) have several examples in my posted R code that you can use. You are *not* required to use my code or my preferred plotting strategies. If you have your own preferred method, that is perfectly fine! Just make it look nice).

$$pmccain \quad N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 pcollege_i + \beta_2 medinc_i + \beta_3 pblack_i + \beta_4 phisp_i + \beta_5 ginicnty_i$$