PUAD 7972: Homework 4

Directions: Download these data, https://davidahughes.files.wordpress.com/2020/04/hw_data.dta_.zip, which are adapted from the 2017 *American Journal of Political Science* article by Monogan, Konisky, and Woods, "Gone with the Wind: Federalism and the Strategic Location of Air Polluters."[1] Using a decision rule of $\alpha = 0.05$ (two-tailed), complete the following prompts. Your work is due by May 1, 2020 (11:59:59 pm).

1. Let the dependent variable represent the total number of major air pollution sites in a given state. Let independent variables include a measure of a state's "green index," a scaled measure environment groups in a given state, the total number of sites built in a given state since 2005, the total number of sites in a given state that are owned by the public, and the log of the raw number of potential air pollution sites in a given states. For each variable, provide summary statistics in a table. Furthermore, provide an alternative hypothesis for each variable.
2. In a regression table, set aside three columns for results (to include coefficients and standard errors). In the first column, present the results from a Poisson regression that includes every variable outlined in Number 1. In the second column, present results from a negative binomial regression. And in the third columns, present results from either a multilevel Poisson or negative binomial (depending upon which is more appropriate given the data) that include random intercepts for each of the four regions in the data.
3. Which set of regression results is the most appropriate given statistical testing? Why? For that model, interpret and present graphically each of the statistically significant results. If the best fitting model is the multilevel model, estimate and present each of the group-level intercepts either textually or graphically and report what they tell us about what affects the number of major air pollution sites in a given state.

Appendix: Codebook for Data

- loc_state: A string variable identifying the state of analysis.
- region: Nominal-level variable for geographic region, "1" for south, "2" for west, "3" for northeast, and "4" for Midwest.
- region_string: String variable indicating a given region.
- gi_tot: Green index based on 67 separate indicators of state environmental policy effort (Hall & Kerr 1991).
- scaledig: Gray and Lowery's environmental interest group data, scaled by the state's Gross State Product.
- total_air: Total number of sites in a given state that are a major source of air pollution.
- total_new: Total number of sites in a given state constructed since the year 2005.
- total_public: Total number of sites in a given state owned by the public.
- site_count: Total number of sites in a given state that could feasible be classified as a major source of air pollution.
- log_site_count: Natural log of the variable, "site_count."

[1] To access the article, navigate here.