Bayesian Analysis of County-wise Demographics in NC for US Presidential Elections

Krishnachaitanya Pullakandam, Venkata Pasumarty, Vikram Patil

MOTIVATION

In 2016, the GOP had a historic victory in the U.S. presidential elections, proving all the polls wrong. Factors such as rural population vote, gender, ethnicity, age, and income groups were some of the contributors which led to this landslide victory. In this study, we aim to draw more detailed insights into the significant determinants that lead to this result, particularly in the state of North Carolina.

DATA

County wise demographic information and voting statistics for all the states was obtained for the study. Data includes demographic information about age, gender, education, ethnicity, total votes secured by GOP, Democratic and other parties.

FEATURE SELECTION: PCA

- The raw data contains highly correlated covariates as seen from the correlation plot.
- **Principal Component Analysis** was used to transform the covariates to linearly uncorrelated variables
- 26 PC's explain 99% of the variance of the 46 covariates. MCMC sampling was performed on the loadings and back transformation was done to obtain the original regression coefficients.

Principal Component Analysis:

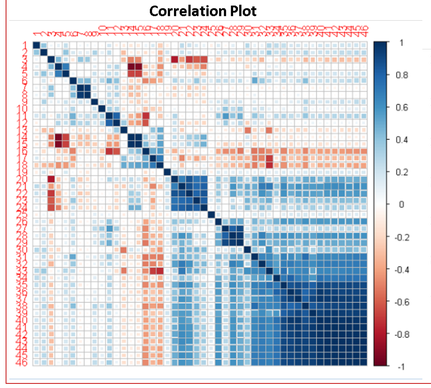
$$S_x = \Gamma D \Gamma' \approx \Gamma_q D_p \Gamma'_q$$

$$X^*_{nxq} \approx X_{nxp} \Gamma_{pxq}$$

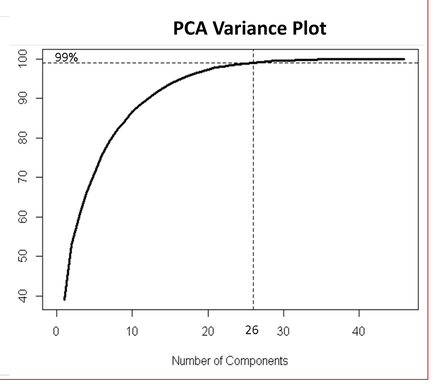
PC Regression: $E(Y) = X^*_{nxq}\, \beta^*_{qx1} = X_{nxp}\, [\Gamma \beta^*]_{px1} = X\beta$

where, S_x = Original covariance matrix
Γ = Eigen vectors
β^* = PCA regression coefficients
β = original Regression coefficients

Correlation Plot



PCA Variance Plot



MODELS

Model 1: Binomial Regression

Likelihood: $\mathbf{Y}_i \sim \text{Binomial}(\mathbf{p}_i, N_i)$
$\text{logit}(\mathbf{p}_i) \sim \alpha_k + \phi * X_{urban} + \Sigma X^*_i\, \beta^*_i$

Prior: $\beta^*_i \sim N(0, 100)$ where # of PC loadings, i = 1,2…..26
$\Phi \sim N(0, 100)$
$\alpha_k \sim N(0, 100)$ where k = 1,2 (2012 and 2016)

- Y_i is # of votes received by GOP in i^{th} county.
- p_i is the probability of GOP's victory over the democrats/others.
- N_i is the total # of votes cast in i^{th} county.
- X is a matrix containing 46 covariates and X_{urban} is urban factor.
- Uninformative Normal priors were used on the parameters.

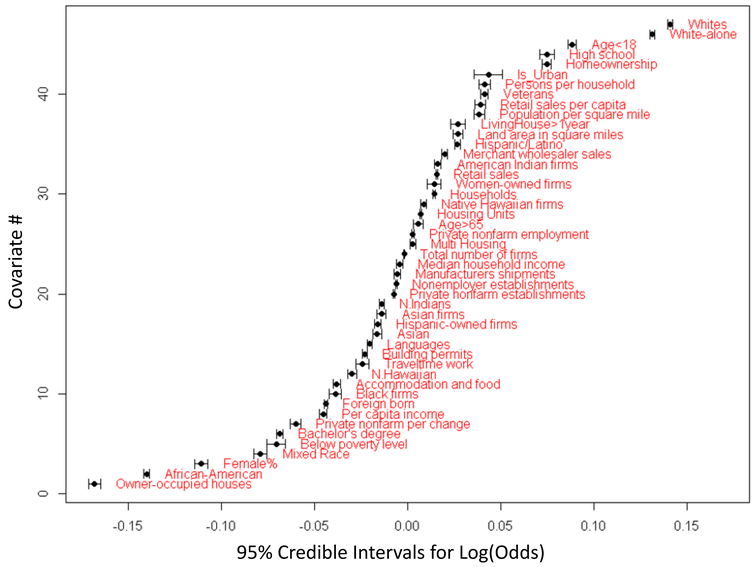
Model 2: Multinomial Regression

Likelihood: $\mathbf{Y}_{i,j} \sim \text{Multinomial}(\mathbf{p}_{ij}, N_i)$
$\mu_{ij} = \alpha + \phi_j * X_{urban} + \Sigma X_i\, \beta^*_{ij}$
$\mathbf{p}_{ij} = e^{\mu ij} / (1 + \Sigma e^{\mu ij})$

Prior: $\beta^*_{ij} \sim N(0, 100)$ where # of PC loadings, i = 1,2…..26; j = 1,2
$\Phi_i \sim N(0, 100)$ where j = 1-Democrat, 2-GOP
$\alpha \sim N(0, 100)$

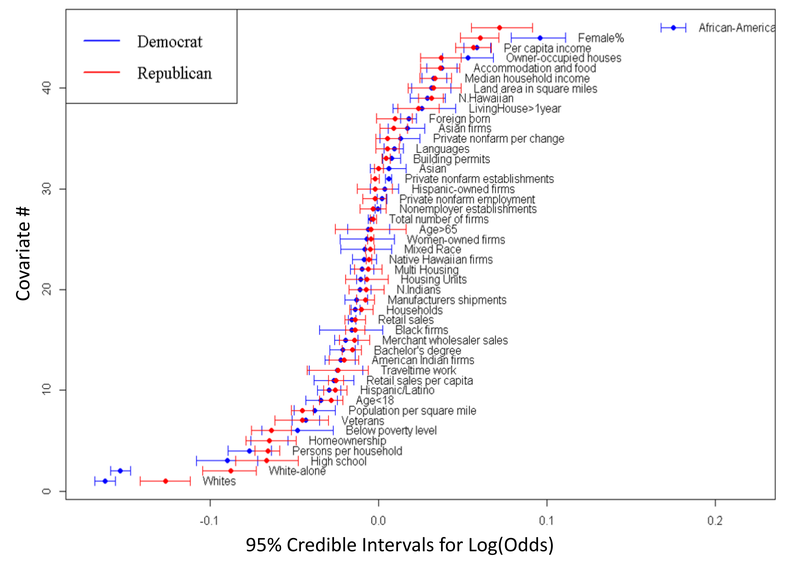
- $Y_{i,j}$ is # of votes received by j^{th} party in i^{th} county.
- $p_{i,j}$ is the probability winning of j^{th} party over others in i^{th} county.
- N_i is the total # of votes cast in i^{th} county.
- X is a matrix containing 46 covariates and X_{urban} is urban factor.
- Uninformative Normal priors were used on the parameters.

BINOMIAL: RESULTS



- Owner occupied houses, African-American population, and female population are some of the factors negatively affecting GOP votes.
- Top factors for higher GOP votes are white population, urban population, population with a high school degree.
- Additional effects can be implemented in further analysis to account for geo-spatial effects.

MULTINOMIAL: RESULTS

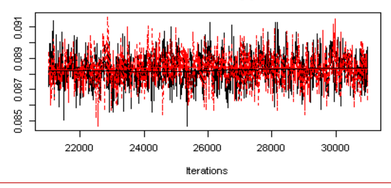


- African-American population, female population are some of the factors negatively affecting GOP votes.
- Top factors for higher GOP votes were white population, urban population, population with a high school degree.
- Many of the factors have over-lapping credible intervals, which tells that those covariates do not have strong party preferences.

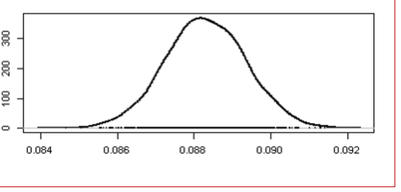
IMPLEMENTATION

- Model was fit in R using: *rjags* package
- 2 Markov chains
- 20,000 burn-in & 10,000 iterations

Trace of beta[1]



Density of beta[1]



REFERENCES

Data Sources:
https://github.com/tonmcg/US_County_Level_Election_Results_08-16
https://www.kaggle.com/benhamner/2016-us-election