Statistical Analysis Plan (Notes)

Assessing the impact of socio-economic factors on Presidential Primary Election voting in the USA in 2016

Aim

What factors affect the probability of republican winning a county?

Outcome

- **Binary outcome**: Republican won or not (logistic regression)

- **Proportion**: Either beta regression model or normal regression model applied to

Missingness

Alaska - Remove as different areas.

Exploratory Analysis

- Maps of binary outcome or proportion republican

- Maps of other key variables

- Scatterplots / boxplots of response vs covariates - linear or non-linear relationships

- Collinearity - Scatterplots between covariates

Variable Reduction

- Choose based on context of variable or PCA to get independent covariates. PCA may be hard to explain.

Modelling

- Logistic Regression either by stepwise or put all variable in model and remove non-significant ones one at a time.

Compare Multiple Models

- Model fit measures - AIC / BIC

- Prediction - RMSE based on 10 fold cross validation. Sensitivity, specificity.

Advanced Chapter

- Model spatial correlation - CARBayes

- Clustering of counties

- Other classification techniques