Quick Memo About Tests We’ve Run:

Klaverns Anova: We ran an anova to test for significant difference in the mean Klaverns per cluster. We found significant difference in means p<.001. We ran a Tukey post-hoc pairwise comparisons test and found that statesfips 22 28 and 37 consistently showed significantly higher numbers of mean klaverns per cluster at p<.01.

Test of Influence: We tested for influence, (dfbetas)/ (2/sqrt(N)). Running regression with and without observations with the most influence. Significance changed but direction of the relationships did not. High influence cases fell on the extremes of our predictor (at X=0 and X=1), this makes sense for the nature of the variable . Since the direction of relationships did not change and the “extremeness” of the values are an artifact of how the variable is calculated, we decided to leave the high influence cases in the regression.

Test of homoskedasticity: We ran vce(hc3) to test for heteroskedasticity. With the hc3 robust standard errors, coefficients and significance were similar to the regression run without the robust standard errors leading us to concluded the errors were relatively homoscedastic and vce(hc3) was not needed.

vce(hc3) uses u 2 j /(1 − hjj ) 2 as suggested by Davidson and MacKinnon (1993 and 2004), who report that this often produces better results when the model is heteroskedastic. vce(hc3) produces confidence intervals that tend to be even more conservative

VIF: Mean VIF was 1.75, highest VIF was 2.49. We concluded no variables needed to be removed from the regression for inflation.

Residuals: we plotted the residuals of our regression and concluded they were homoscedastic reiterating our conclusions above with robust standard error.

Non-linearity: we plotted all predictors to look for non-linearities. We spotted some non-linearity with the total population which was positively skewed. When logged, and re-entered into the regression the results were very similar. We decided to keep the non-transformed totpop variable to keep interpretation simple for the reader. All tests of significance for coefficients and for model were similar with the linear and logged population variable.