

POL 168: Latin@ Politics

Fall 2008

Problem Set 4: Analysis of Aggregate Data from the 2008 Election

For this assignment, I want you to do some primary analysis of aggregate election statistics I have compiled for you. This assignment is due Monday, Nov. 24. This is plenty of time if you start on this NOW. If you wait, you will not finish it in sufficient time.

Posted on the website is an Excel spreadsheet. If you use a Mac, you can download openoffice from the Apple website to access Excel spreadsheets. The data in the spreadsheet gives the vote share (and proportions) for Propositions 4 and 8 for each California county from the 2008 election. Additionally, the racial/ethnic characteristics for each county is given. There are several columns in the spreadsheet that are clearly labeled. They are:

Col 1: Name of county

Cols 2-3: Votes FOR and Votes AGAINST Prop. 8

Col 4: Percentage of vote in favor of Prop. 8

Cols 5-6: Votes FOR and Votes AGAINST Prop. 4

Col 7: Percentage of vote in favor of Prop 4

Cols 8-11: Percentage of: black, Hispanic, white n.h., Asian

Please use the available data to answer the following questions, do the following tasks.

1. Provide a scatterplot of the percentage of vote on Proposition 8 versus Proposition 4. Describe the basic features of this plot (i.e. do counties with high votes on one issue have high votes on the other issue?) Which counties are high? Which low? Are there any patterns you note here? Note: first sort the data by one of the Propositions. This will make the graph easier to interpret. (20 points)

2. Compute a Pearson correlation coefficient for each racial/ethnic group on the percentage vote for each ballot issue. In Excel, this requires a cell like `=PEARSON(column1, column2)` where column 1 corresponds to the vote proportion (on 4 or 8) and column 2 corresponds to one of the columns for race/ethnic breakdown. For example, if I want the correlation between percent Asian and vote on Prop 2, the Excel code is: `=PEARSON(D2:D59, K2:K59)`.

A Pearson's r ranges from -1 to 1 , where 1 means a perfectly positive association and -1 means a perfectly negative association. Correlations for data like this above, say $.20$ or below $-.20$ are generally considered large to moderately large. Thus, if I found the correlation between percentage black in the county and vote percentage on Prop. 8 were $.25$ (it is not), I

would infer that counties having a larger African-American population tended to be associated with higher vote percentages on Prop. 8. Neatly tabulate the correlation coefficients. (20 points)

3. For each correlation, provide an interpretation of the statistic. Is it large? Is it small? What do the data suggest in particular about the correlation within the Hispanic and African-American communities? Based on the analysis you did in Homework 3, do these correlations seem consistent with those drawn in that assignment with respect to the Latino population. (30 points)

4. In just looking at these correlations, would it be fair to assume that “most Latinos” in counties where the vote share is high on Props. 4 and 8, voted “for” the ballot measures? That is, can these data be reliably used to make conclusions about individual Latino voters? Why or why not? (10 points)

5. Besides racial/ethnic voting patterns on these ballot issues, what other factors may contribute to high/low voting percentages in a county on these issues? Please explain. (10 points)