Empirical Project

Overview

Imagine that you have been hired as a political consultant by the Republican Party for the next presidential election. California is a difficult state and it has been carried by democrats in the last few elections. Your goal is to analyze the socio-economic background that explains this tilt. Part of your job is to identify what factors seem to explain the democrat vote. Based on your findings, you will be giving advice on how to allocate the electoral budget for California. In addition, you may also want to give advice to the California republicans as to what type of initiatives are more likely to earn votes for the next election – voting in favor of certain budgetary initiatives may be perceived warmly by the electorate. How careful you are in your report will translate into a higher consulting fee – i.e. grade. Make sure that all your recommendations and comments are entirely and dispassionately based on the data – any good defense lawyer will tell you, you need not like your client to do a good job!

Instructions

• You will be asked to work in teams of two people at most. If you cannot find a teammate, you can choose between doing the report alone or to be allocated to another team at random.

• The final report should be no more than 10 pages long, including graphs, tables and the like. The first page should contain the executive summary: 99% of the time, this is the only part of a report that ever gets read so be sure to include your punch-line!

• In the main text of the report, be sure to economize on the space you use to report regression output. Direct EViews output is not acceptable. Your regression work should be reported in tables, such as is done in your book, for example, table 5-2, p. 182 or you can report individual regressions in the text, such as is done in page 181.

• If you choose to do hypotheses tests, please report these in tables on simply in the text, by explaining the purpose of the test, the null hypothesis, and the outcome (with the p-value in parenthesis if you prefer). For example: … to test whether ethnicity is an important determinant of party vote, we conducted a joint test of the null hypothesis that all the coefficients for the ethnicity variables were zero. The results support the null with a p-value of 0.45.”
• Graphs are great but only if they offer meaningful information, otherwise they are a waste of space. If you include graphs, be sure to label every component: the axis, the units, what variables are being displayed, etc. Include this information in the graph and use a note at the foot of the graph to explain its purpose.

• Both graphs and tables should be self-contained: if someone pulled your report and just read the graphs and the tables, the person should be able to understand them without having to read the text.

• Be sure to structure your report appropriately. Your introduction should be no more than two paragraphs. The main goal is to explain what’s coming, how you are going to achieve your goals, and to give a preview of the most important findings.

• The next section should give an overview of the data (summary statistics and the like). There are many variables you may wish to transform (for example, percentages are more easily interpreted than absolute figures). Be sure to report the summary of the final transformed data that you will use in the analysis only.

• The next section should contain the bulk of your empirical analysis. I expect that you will be able to identify two or three main issues and that you will run a few regressions to try to come with answers to these two or three issues – do not go overboard and try to run every regression that crosses your mind. Use a couple of counties (such as Los Angeles and San Francisco to give contrasting examples that illustrate your points.

• Finally, you should end the paper with your conclusions. This section should summarize your findings, making references back to the empirical section to back your arguments. 75% of the quality of a report rests with the ability to write a good introduction and a good conclusion. Most people read these sections, then look at tables and graphs and never read the rest.

The Data

Data for this assignment is in the excel file “data.xls” It contains data broken down by county on the following variables:

• POPULATION 1990
• POPULATION BY RACE, WHITE 1990 (1C)
• POPULATION BY RACE, BLACK 1990 (1C)
• POPULATION BY RACE, AMERICAN INDIAN, ESKIMO, OR ALEUT 1990
• POPULATION BY RACE, ASIAN OR PACIFIC ISLANDER 1990
• HISPANIC ORIGIN POPULATION (MAY BE OF ANY RACE) 1990
• MALES PER 100 FEMALES 1990
• COMMUNITY HOSPITAL BEDS PER 100,000 POPULATION 1991
• SOCIAL SECURITY PROGRAM BENEFICIARIES PER 1,000 POPULATION 1990
• SERIOUS CRIMES PER 100,000 POPULATION 1991
• PERSONS 25 YEARS AND OVER, PERCENT HIGH SCHOOL GRADUATE OR HIGHER 1990
• PERSONS 25 YEARS AND OVER, PERCENT WITH BACHELOR'S DEGREE OR HIGHER 1990
• MEDIAN FAMILY MONEY INCOME 1989
• PERCENT OF FAMILIES WITH INCOME BELOW POVERTY LEVEL 1989
• OWNER-OCCUPIED HOUSING UNITS, PERCENT OF TOTAL OCCUPIED 1990 (1C)
• VEHICLES AVAILABLE, AVERAGE PER HOUSEHOLD 1990
• CIVILIAN LABOR FORCE, UNEMPLOYMENT RATE 1991
• EARNINGS, PERCENT GOODS-RELATED, MANUFACTURING 1990
• EARNINGS, PERCENT SERVICE-RELATED, TOTAL 1990
• EARNINGS, PERCENT SERVICE-RELATED, GOVERNMENT 1990
• FARM EARNINGS, PERCENT OF TOTAL EARNINGS 1990
• FEDERAL FUNDS AND GRANTS, PER CAPITA EXPENDITURES 1992
• FEDERAL GOVERNMENT CIVILIAN EMPLOYMENT PER 10,000 POPULATION 1990
• STATE AND LOCAL GOVERNMENT EMPLOYMENT PER 10,000 POPULATION 1990
• VOTE CAST FOR PRESIDENT, PERCENT DEMOCRATIC 1992 (COPYRIGHT)
• VOTE CAST FOR PRESIDENT, PERCENT REPUBLICAN 1992 (COPYRIGHT)
• VOTE CAST FOR PRESIDENT, PERCENT PEROT 1992 (COPYRIGHT)
You may not need all the variables for your regressions. Also, you may want to construct variables differently than how they are reported. For example, calculate the percent of Asians in the county rather than just the actual number (or perhaps just the ratio relative to the population in that county).

Beware of multicollinearity! A number of variables are percentages that will add up to 100% if you include all the variables mindlessly.