Case Study 3 - A Study of Demographic Differences against Party Preference

Elliot Pickens & Dean Gladish

May 22, 2018

Introduction

In this modern world, party preference is typically thought to be associated with the demographics and geography of a populace. It is of interest to politicians and the media alike to determine the extent of such correlation in order to understand which groups are most likely to vote for the party. Our case study, which uses data collected from U.S. adults from the 1980 and 2000 elections respectively as part of the National Election Studies project, is an investigation into the matter that allows us to model party preference using the logistic regression model. Specifically, we aim to address whether gender, regional, and union differences play a part in party preference over time.

Data

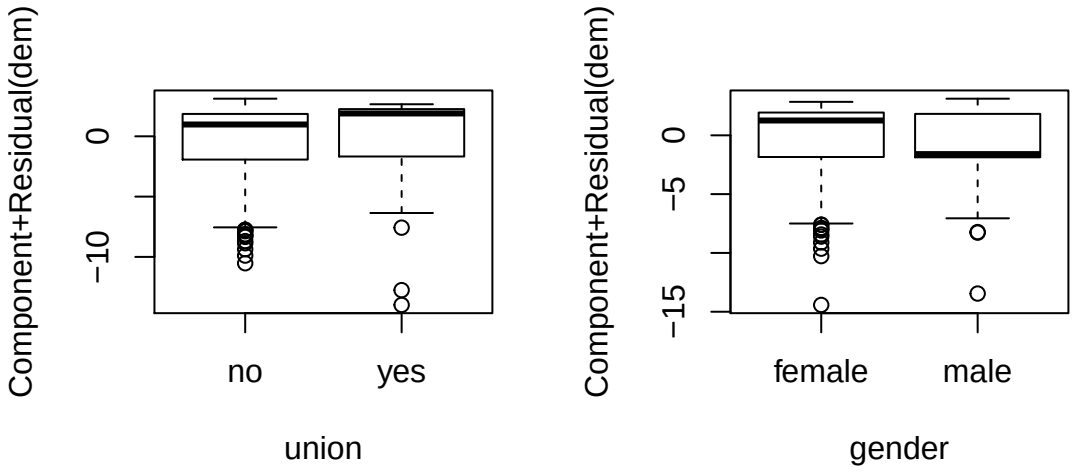
The dataset that we analyzed consists of a binary indicator variable indicating Democratic Party preference as well as numerous other categorical variables corresponding to factors such as year, age, gender, race, region, income, unionized and educational status. The three explanatory variables that we focus on are *gender*, *region*, and *union*.

The following table gives our estimates of the important aspects (coefficients, etc.) of our model:

Table 1: Important Coefficients of our Logistic Regression Model

	Estimate	Standard error	z value	P-value
intercept	1.648114	0.221279	7.448	9.47e-14
raceother	-1.600862	0.227855	-7.026	2.13e-12
racewhite	-1.852649	0.182748	-10.138	2e-16
unionyes	0.692776	0.116536	5.945	2.77e-09
incomemiddle 1/3	-0.265841	0.113745	-2.337	0.01943
incomeupper 1/3	-0.491922	0.114198	-4.308	1.65e-05
age	0.007781	0.002699	2.883	0.00395
gendermale	-0.258311	0.090486	-2.855	0.00431

The following set of plots represent what is essentially the relationship between our binary model and the data for the years 1980 and 2000.



Through exploratory data analysis of significance and association, we found that interaction variables gave us a closer fit to the data.

Results:

Using the AIC criterion we obtained the following model:

$$\widehat{Y}_i\{dem\} = \beta_0 + \beta_1 raceother + \beta_2 racewhite + \beta_3 unionyes + \beta_4 incomemiddle\ 1/3 + \beta_5 incomeupper\ 1/3 + \beta_6 age + \beta_7 gendermale$$

The model shown above can be interpreted as follows:

Discussion: