

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*. We came up with the model

$$\hat{Y}_i\{dem|\} = \beta_0 + \beta_1 Gender$$

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	-1.576601	9.492598	-0.166	0.86809
gendermale	-0.25969	0.090771	-2.861	0.00422
regionNE	0.10458	0.134548	0.777	0.437
regionS	0.008454	0.11912	0.071	0.94342
regionW	0.175367	0.134395	1.305	0.19194
unionyes	0.682129	0.120305	5.67	1.43e-08

The following set of scatterplots represent what is essentially the relationship between our estimated 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: