Finding Supporters

Political Predictive Analytics

Using Logistic Regression

Multivariate Solutions

What is ‘Logistic Regression’?

- **In a political application, logistic regression can describe the outcome of whether a person will ‘vote’ or ‘not vote’; or ‘Vote for my candidate’, ‘Not Vote for my candidate’.**
- **Logistic regression should be used in situations when a campaign would like to ascertain which voters are:**
  - **Likely Voters**
  - **Inclined to vote for their candidate**
  - **‘Undecided’**
  - **Donor possibility**
- **Logistic regression in a nutshell:**
  - Logistic regression is used for prediction of the probability of occurrence of an event by fitting data to a logistic curve.
  - Logistic regression makes use of several predictor variables that may be either numerical or categorical.

Example: Finding the ‘Liz Anderson’ Supporter

- For example, in a political race, logistic regression could be used to analyze the factors that determine whether an individual should be targeted with strategic communications; in the same sense as a ‘swing’ state in president elections.
- In our example, our candidate, Liz Anderson, is running for Congress.
- Let us define a variable ‘Outcome – Vote for Liz Anderson’
  - Outcome = 1 If 'The Individual Is Likely to vote for Liz’
    = 0 If 'The Individual Is Not Likely To Vote for Liz’.

- The outcome takes only *two possible values*.

Variables in the Model

- **These are the variables in the ‘Anderson for Congress’ model**
  - Gender-Male=1,Female=2
  - Married
  - Own home
  - Urban=1/Rural=0
  - Working women in Household
  - Completed college
  - Credit card debt

Regression Output

Vote for 'Liz' Logistic Regression			
Regression Output	**Regression Beta**	**Sig.**	**Odds Ratio**
Gender 1=Male 2=Female	2.670	0.029	14.440
Marital Status 1=Married/0=Not Married	2.095	0.015	8.126
College Graduate 1=Yes, 0=No	1.800	0.120	6.050
Working Woman in House 1=Yes, 0=No	1.762	0.038	5.826
Urban 1=Urban, 0=Not urban	1.649	0.024	5.202
Housing 1=Own, 0=Rent	0.800	0.506	2.226
Credit Card Debt 1=Yes, 0=No	-0.860	0.252	0.423

Shading indicated statistical significance

This slide is descriptive, and shows which of the variables are most influential in determining whether a voter is a ‘Liz for Congress’ supporter.

The campaign can now determine, by examining this table, that educated married, working woman who live in urban areas are a key demographic to target in the days leading up to the election. These should be targets for both **Strategic Communications** and **Get Out the Vote** field operations.

Odds-Ratio

- **When a respondent's choices are set within the regression model, an 'odds-ratio' for each respondent is created using the formula of $1/(1+e^{-z})$.**
  - 'Z' is the outcome of the regression equation once all the questions are input.
- A simulator can be used to classify individuals based on demographic data or a survey screen.
- Two examples follow:

Example One

Urban College-Educated Working Woman

Vote for 'Liz' Logistic Regression			
Regression Output	**Answer**	**Regression Beta**	**Sum (b*c)**
Gender 1=Male 2=Female	2	2.670	5.340
Marital Status 1=Married/0=Not Married	1	2.095	2.095
College Graduate 1=Yes, 0=No	1	1.800	1.800
Working Woman in House 1=Yes, 0=No	1	1.762	1.762
Urban 1=Urban, 0=Not urban	1	1.649	1.649
Housing 1=Own, 0=Rent	1	0.800	0.800
Credit Card Debt 1=Yes, 0=No	0	-0.860	0.000
Sum			**3.446**

Odds Ratio ($1/(1+e^{-z})$)	**0.97**
(Chances to 'Vote for Liz')	**97%**

Using the logistic output, the chances of a college educated, urban, working, married woman who owns her own home and has no credit card debt to be a supporter shown above. The logistic regression formula ($1/(1=e^{-z})$) gives this woman an odd-ratio (in purple) of .97, or 97% chance that she will vote for Liz.

Example Two

Urban College-Educated Working Man

Vote for 'Liz' Logistic Regression			
Regression Output	**Answer**	**Regression Beta**	**Sum (b*c)**
Gender 1=Male 2=Female	1	2.670	2.670
Marital Status 1=Married/0=Not Married	1	2.095	2.095
College Graduate 1=Yes, 0=No	1	1.800	1.800
Working Woman in House 1=Yes, 0=No	1	1.762	1.762
Urban 1=Urban, 0=Not urban	1	1.649	1.649
Housing 1=Own, 0=Rent	1	0.800	0.800
Credit Card Debt 1=Yes, 0=No	0	-0.860	0.000
Sum			**0.776**

Odds Ratio ($1/(1+e^{-z})$	**0.68**
(Chances to 'Vote for Liz')	**68%**

Using the logistic output, the chances of a college educated, urban, working, married MAN who owns his own home and has no credit card debt to be a 'Liz' supporter is shown above. The logistic regression formula ($1/(1=e^{-z})$ gives this man an odd-ratio (in purple) of .68, or 68% chance that he will vote for Liz (down from 97% for a women with the same demographic characteristics).