Assignment 2 - Obama & Happiness - Discrete Choice Modelling

Marc Sparhuber

Table of contents

Data & Descriptives

Table 1: Descriptives split by voting for Obama

		voted Obama				didn't vote Obama				combined			
		Mean	SD	N	Percent	Mean	SD	N	Percent	Mean	SD	N	Percent
Happiness	Very Happy			269	25.43			221	34.80			490	28.94
	Pretty Happy			622	58.79			355	55.91			977	57.71
	Not Too Happy			167	15.78			59	9.29			226	13.35
Bachelor or Graduate degrees of parents	0			787	74.39			475	74.80			1262	74.54
	1 or more			271	25.61			160	25.20			431	25.46
Sex	Male			420	39.70			310	48.82			730	43.12
	Female			638	60.30			325	51.18			963	56.88
Age		52.32	16.76	1058	100.00	56.00	16.31	635	100.00	53.70	16.68	1693	100.00
All				1058	100.00			635	100.00			1693	100.00

Comments: General Social Survey data from the socviz R package.

The following analysis tests the hypothesis that individuals who were more unhappy in 2016, were more likely to have voted Obama in 2012. This is expected to be the case because in 2016 individuals who had voted Democrat in 2012 were now faced with Donald Trump as their future president, in some cases leading to protests (SOURCE). To test this hypothesis, a subset of data from the General Social Survey, conducted by the National Opinion Research at the University of Chicago and provided as part of the socviz package in R is used.

Table 1 displays descriptive statistics for all variables used for modelling, split by having voted Obama or not, as well as the entire sample. It shows number of respondents and percent within its group and variable for categorical variables and extends this by means and SDs for continuous variables. Of the 1693 individuals included in the analyses, 1058 (62.49%) voted for Obama, while the 635 (37.51%) did not. It is important to note that this indicates a slightly skewed sample, as Obama received 51.1% of the popular vote in 2012 (SOURCE). As the main independent variable, Happiness is split into the categories 'Not Too Happy', 'Pretty Happy' and, 'Very Happy'. Across the entire sample, more than half of the respondents reported being 'Pretty Happy' (57.71%), 28.94% reported being 'Very Happy', while 13.35% considered themselves 'Not Too Happy'. Relevant to the hypothesis, while only 9.29% of those who did not vote for Obama reported being 'Not Too Happy', 15.78% of those who voted Obama did. While the responses for being 'Pretty Happy' were similar across respondents who had voted Obama or not - 58.79% and 55.91%, respectively - noticeably fewer of the Obama voters considered themselves 'Very Happy' (25.43%) than the others (34.80%).

Further categorical variables describing the sample include whether the respondent had at least one parent with a higher education degree, sex, and age. Interestingly, there seems to be no connection between having voted Obama and higher education, with 75% of this sample having no college educated parent and this percentage not varying across having voted Obama or not. Looking at the sex variable, although

56.88% of respondents were female, voting seemingly split along gender lines with women outweighing men among Obama voters with 60.30% and men being more represented among those who did not vote for Obama with 48.82%. Regarding the age of respondents, the sample mean is 54 years. This varies only slightly across the two conditions, with Obama voters being slightly younger average (52) than non-Obama voters (56). The standard deviation stays at at around 16 in all conditions.

Model Estimaton & Odd Ratios

Table 2: Voting for Obama. Logistic probability models

	Model 1	Model 2	Model 3
(Intercept)	2.831***	2.803***	5.142***
	[2.117, 3.838]	[2.091, 3.809]	[3.189, 8.382]
happyVery Happy	0.430***	0.427***	0.431***
	[0.303, 0.605]	[0.300, 0.602]	[0.302, 0.609]
happyPretty Happy	0.619**	0.615**	0.614**
	[0.445, 0.852]	[0.442, 0.847]	[0.440, 0.849]
parent_degree1		1.061	0.924
		[0.845, 1.335]	[0.729, 1.174]
age			0.985***
			[0.979, 0.992]
sexFemale			1.483***
			[1.212, 1.815]
Num.Obs.	1693	1693	1693
AIC	2220.6	2222.3	2192.1
BIC	2236.9	2244.1	2224.7
Log.Lik.	−1107.301	−1107.172	−1090.037

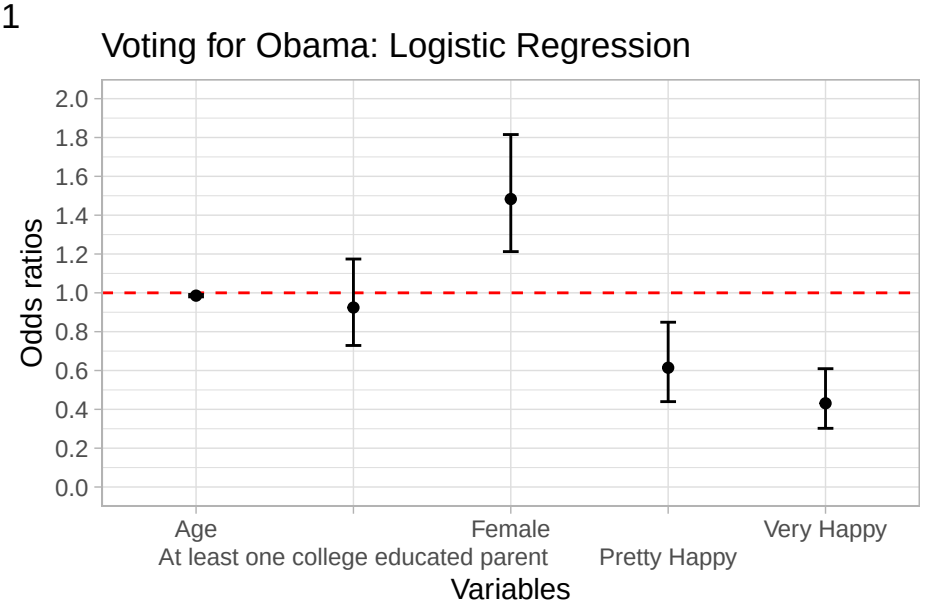
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Source: General Social Survey data from the socviz R package.

Comments: The reference cateogory for happy is 'Not Too Happy'.

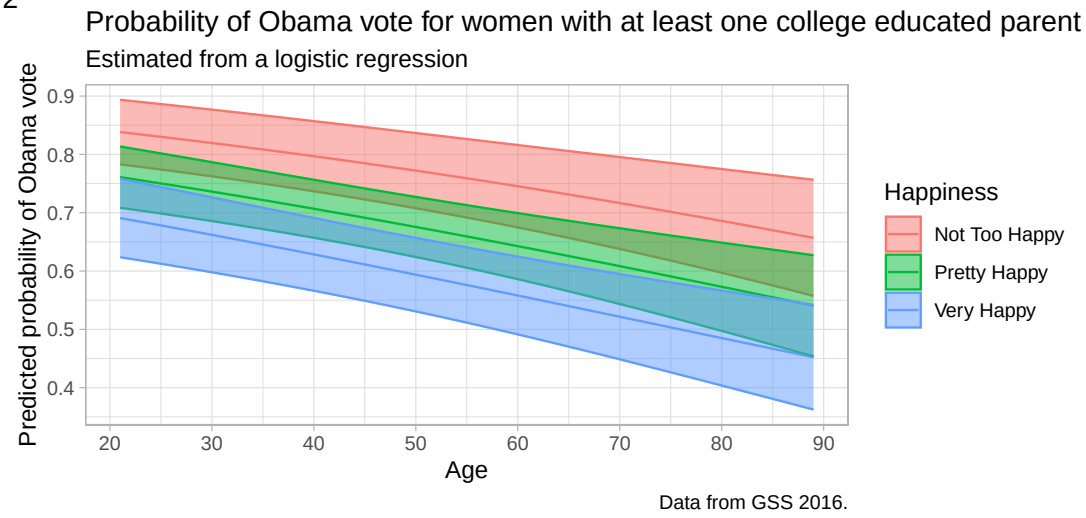
Table 2 shows the results of three logistic probability models. Each model contains the full sample of 1693 observations and presents the coefficients as odds ratios with corresponding confidence intervals in square brackets. Model 1 only uses Happiness as

Figure 1



Predicted Probabilities

Figure 2



Model Fit

Table 3: Voting for Obama: Model fit statistics

	Model 1	Model 2	Model 3
Nagelkerke's pseudo-R2	0.02044	0.02064	0.04754
Share of correct predictions	0.62493	0.62493	0.62788
Likelihood Ratio		0.61086	3.6181e-08

Note: A prediction is considered correct when its probability is greater than 0.5. The Likelihood ratio is always calculated with the model to the left. Data from GSS 2016.

Conclusion