

Logistic Regression

Goal: Use the Jeffres National Community Study data set to predict the probability of a person circulating or signing a petition.

I. The Model

Variables: In Logistic Regression you must choose a DV that is a binary or can be dummy coded. Running frequency tests lead us to select if a person had ever circulated or signed a petition. Demographics was the obvious name given to the first block of predictors, since we included education level, race (dummy coded white= 1; all others= 0), age, and gender (dummy coded femaleness= 1; male= 0) as our variables. Block two was named community participation; and we included “I generally discuss political candidates and issues with neighbors at election time,” “have you worked with others in your community to solve some community problem in the past couple years,” and “attended meetings of your town or city council,” because these variables dealt with participating in community activities. We titled Block three sense of government security, and put the variables “public officials in my community seem receptive to views of residents,” and “other than voting, people like me have little influence over local government actions.” Lastly, Block four was called political involvement and had the variables membership in political clubs or organizations, worn a button or put a sticker on your car, voted in the 2004 presidential election, and contributed money to a party or candidate put in since they show political activeness.

II. Running SPSS

SPSS allows a hierarchical ordering of the blocks, where contribution increases downward, and automatically dummy codes the dependent variable for you...

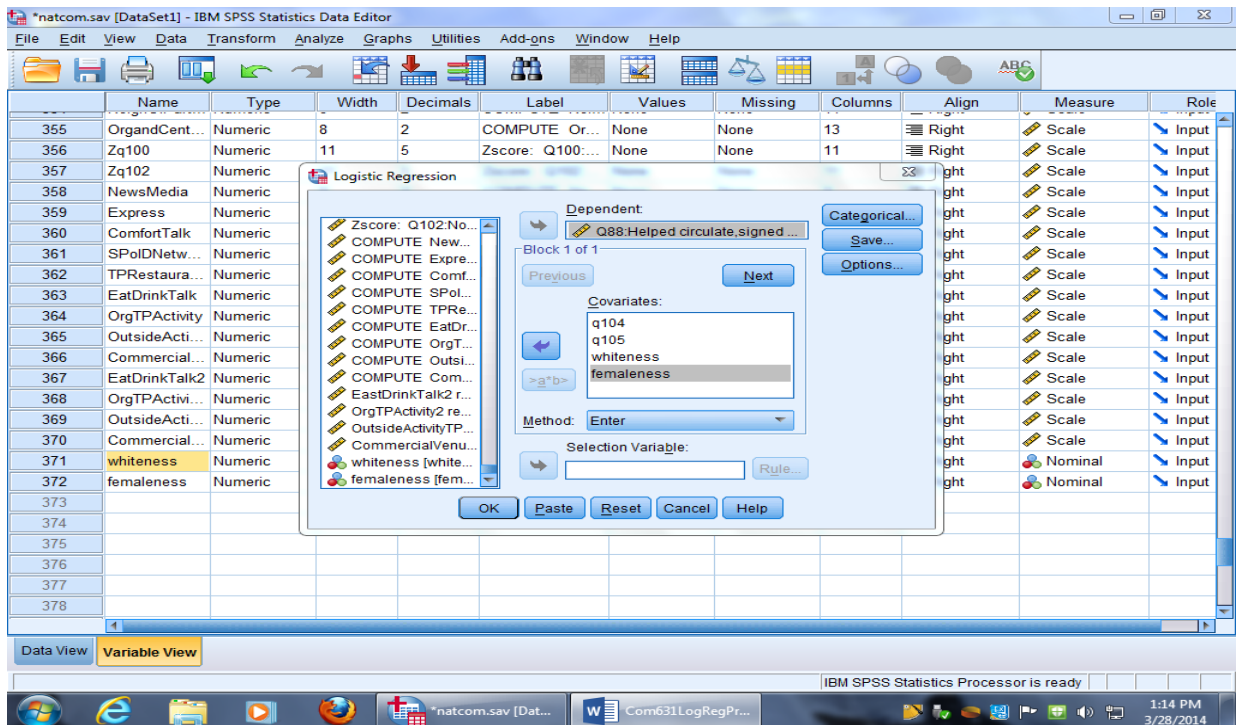
Steps: Forced entry method

CLICK: ANALYZE

REGRESSION

BINARY LOGISTIC

DEPENDENT: Insert dependent variable



COVARIATES: Insert independent variables from Block 1

CLICK: NEXT

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a list of variables with columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. A dialog box titled "Logistic Regression" is open, showing the following configuration:

- Dependent:** Q88:Helped circulate, signed...
- Covariates:** q104, q105, whiteness, femaleness
- Method:** Enter
- Selection Variable:** (empty)

The dialog box also includes buttons for "Categorical...", "Save...", "Options...", "OK", "Paste", "Reset", "Cancel", and "Help".

The screenshot shows the IBM SPSS Statistics Data Editor interface. The main window displays a list of variables with columns for Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. A dialog box titled "Logistic Regression" is open, showing the following configuration:

- Dependent:** Q88:Helped circulate, signed petf...
- Covariates:** q24, q52, q83
- Method:** Enter
- Selection Variable:** (empty)

The dialog box also includes buttons for "Categorical...", "Save...", "Options...", "OK", "Paste", "Reset", "Cancel", and "Help".

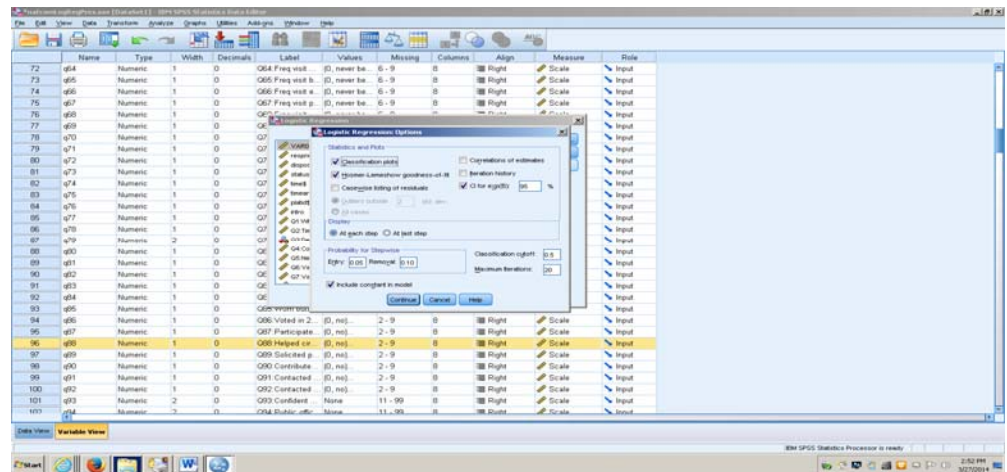
COVARIATES: Insert independent variables from Block 2

The screenshot shows the IBM SPSS Statistics Data Editor with a Logistic Regression dialog box open. The dialog is for Block 4 of 4. The dependent variable is 'Q88: Helped circulate signed pett...'. The method is set to 'Enter'. The covariates listed are 'Q46', 'Q85', 'Q86', and 'Q80'. The dialog box has buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. The background shows a list of variables with their names, types, widths, decimals, labels, values, missing values, columns, align, measure, and role.

The screenshot shows the IBM SPSS Statistics Data Editor with a Logistic Regression dialog box open. The dialog is for Block 3 of 4. The dependent variable is 'Q88: Helped circulate signed pett...'. The method is set to 'Enter'. The covariates listed are 'Q23' and 'Q22'. The dialog box has buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help'. The background shows a list of variables with their names, types, widths, decimals, labels, values, missing values, columns, align, measure, and role.

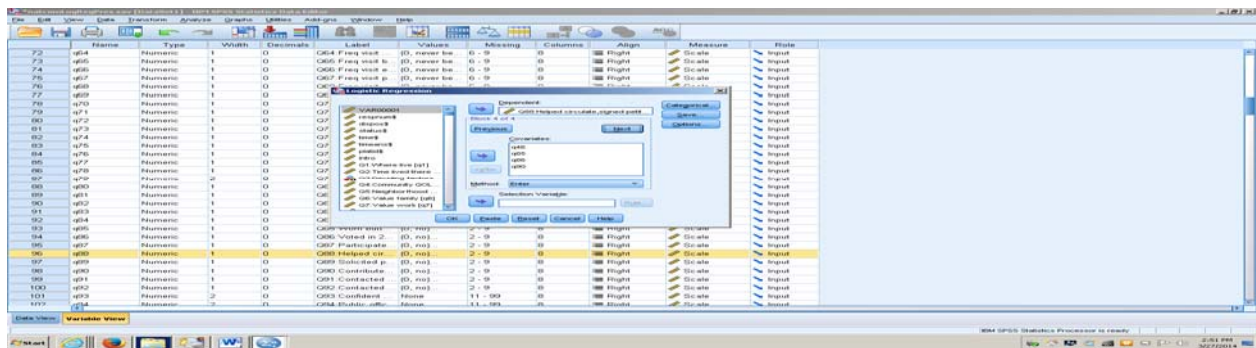
REPEAT: Do same as above for Block 3 and Block 4 respectively

CLICK: OPTIONS



SELECT (Classification Plots, Hosmer-Lemeshow, and CI for exp(B) at 95%)

CLICK: CONTINUE



CLICK: OK

III. SPSS Output

```

LOGISTIC REGRESSION VARIABLES q88
  /METHOD=ENTER q104 q105 whiteness femaleness
  /METHOD=ENTER q24 q52 q83
  /METHOD=ENTER q23 q32
  /METHOD=ENTER q46 q85 q86 q90
  /CLASSPLOT
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

```

Logistic Regression

Notes

Output Created		28-Mar-2014 13:15:46
Comments		
Input	Data	C:\Users\gaoffice\AppData\Local\Temp atcom.sav
	Active Dataset	DataSet1
	File Label	CP05
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	477
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing

Syntax		LOGISTIC REGRESSION VARIABLES q88 /METHOD=ENTER q104 q105 whiteness femaleness /METHOD=ENTER q24 q52 q83 /METHOD=ENTER q23 q32 /METHOD=ENTER q46 q85 q86 q90 /CLASSPLOT /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	00 00:00:00.031
	Elapsed Time	00 00:00:00.063

[DataSet1] C:\Users\gaoffice\AppData\Local\Temp\natcom.sav

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	397	83.2
	Missing Cases	80	16.8
	Total	477	100.0
Unselected Cases		0	.0
Total		477	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
no	0
yes	1

Block 0: Political Action**Classification Table^{a,b}**

Observed			Predicted		Percentage Correct
			Q88:Helped circulate,signed petition		
			no	yes	
Step 0	Q88:Helped circulate,signed petition	no yes	0 0	184 213	.0 100.0
Overall Percentage					53.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.146	.101	2.115	1	.146	1.158

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	q104	.243	1	.622
		q105	8.046	1	.005
		whiteness	.638	1	.425
		femaleness	2.344	1	.126
	Overall Statistics		11.096	4	.026

Block 1: Demographics**Omnibus Tests of Model Coefficients**

		Chi-square	df	Sig.
Step 1	Step	11.226	4	.024
	Block	11.226	4	.024
	Model	11.226	4	.024

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	537.013 ^a	.028	.037

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	9.522	8	.300

Contingency Table for Hosmer and Lemeshow Test

		Q88:Helped circulate,signed petition = no		Q88:Helped circulate,signed petition = yes		Total
		Observed	Expected	Observed	Expected	
Step 1	1	29	24.390	11	15.610	40
	2	19	23.210	23	18.790	42
	3	20	22.754	24	21.246	44
	4	20	20.456	21	20.544	41
	5	16	18.498	24	21.502	40
	6	23	19.265	20	23.735	43
	7	17	15.935	21	22.065	38
	8	12	15.763	28	24.237	40
	9	18	14.657	23	26.343	41
	10	10	9.072	18	18.928	28

Classification Table^a

Observed			Predicted		
			Q88:Helped circulate,signed petition		Percentage Correct
			no	yes	
Step 1	Q88:Helped circulate,signed petition	no	79	105	42.9
		yes	70	143	67.1
	Overall Percentage				55.9

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	q104	-.022	.063	.116	1	.734	.979	.864	1.108
	q105	.223	.078	8.093	1	.004	1.249	1.072	1.456
	whiteness	-.210	.247	.726	1	.394	.810	.499	1.315
	femaleness	-.280	.205	1.872	1	.171	.755	.506	1.129
	Constant	-.363	.441	.678	1	.410	.695		

a. Variable(s) entered on step 1: q104, q105, whiteness, femaleness.

Block 2: Community Participation

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	29.481	3	.000
	Block	29.481	3	.000
	Model	40.707	7	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	507.532 ^a	.097	.130

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.761	8	.674

Contingency Table for Hosmer and Lemeshow Test

		Q88:Helped circulate,signed petition =		Q88:Helped circulate,signed petition =		Total
		no		yes		
		Observed	Expected	Observed	Expected	
Step 1	1	32	28.209	8	11.791	40
	2	23	25.687	17	14.313	40
	3	22	23.833	18	16.167	40
	4	22	21.864	18	18.136	40
	5	16	19.850	24	20.150	40
	6	21	17.596	19	22.404	40
	7	16	15.467	24	24.533	40
	8	14	13.333	26	26.667	40
	9	10	10.818	31	30.182	41
	10	8	7.343	28	28.657	36

Classification Table^a

Observed			Predicted		Percentage Correct
			Q88:Helped circulate,signed petition		
			no	yes	
Step 1	Q88:Helped circulate,signed petition	no	106	78	57.6
		yes	71	142	66.7
Overall Percentage					62.5

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	q104	-.025	.066	.144	1	.705	.975	.856	1.111
	q105	.155	.082	3.576	1	.059	1.167	.994	1.370
	whiteness	-.206	.255	.657	1	.418	.814	.494	1.340
	femaleness	-.113	.215	.276	1	.599	.893	.586	1.362
	q24	.074	.033	5.053	1	.025	1.077	1.010	1.149
	q52	.808	.233	12.037	1	.001	2.244	1.421	3.543
	q83	.356	.245	2.110	1	.146	1.428	.883	2.309
	Constant	-.931	.470	3.919	1	.048	.394		

a. Variable(s) entered on step 1: q24, q52, q83.

Block 3: Sense of Government Security

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	2.243	2	.326
	Block	2.243	2	.326
	Model	42.949	9	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	505.289 ^a	.103	.137

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	12.591	8	.127

Contingency Table for Hosmer and Lemeshow Test

		Q88:Helped circulate,signed petition =		Q88:Helped circulate,signed petition =		Total
		no		yes		
		Observed	Expected	Observed	Expected	
Step 1	1	32	28.592	8	11.408	40
	2	19	25.897	21	14.103	40
	3	25	23.800	15	16.200	40
	4	25	21.846	15	18.154	40
	5	16	19.894	24	20.106	40
	6	20	17.851	20	22.149	40
	7	19	15.035	21	24.965	40
	8	12	13.250	28	26.750	40
	9	11	10.489	29	29.511	40
	10	5	7.347	32	29.653	37

Classification Table^a

Observed			Predicted		Percentage Correct
			Q88:Helped circulate,signed petition		
			no	yes	
Step 1	Q88:Helped circulate,signed petition	no	108	76	58.7
		yes	70	143	67.1
	Overall Percentage				63.2

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	q104	-.006	.068	.007	1	.934	.994	.870	1.136
	q105	.157	.084	3.470	1	.062	1.170	.992	1.380
	whiteness	-.217	.256	.721	1	.396	.805	.488	1.328
	femaleness	-.088	.217	.165	1	.685	.916	.598	1.401
	q24	.082	.034	5.889	1	.015	1.085	1.016	1.160
	q52	.781	.235	11.040	1	.001	2.183	1.377	3.461
	q83	.354	.247	2.047	1	.153	1.424	.877	2.312
	q23	-.051	.039	1.721	1	.190	.950	.880	1.026
	q32	-.036	.035	1.030	1	.310	.965	.900	1.034
	Constant	-.566	.550	1.059	1	.304	.568		

a. Variable(s) entered on step 1: q23, q32.

Block 4: Method = Enter**Omnibus Tests of Model Coefficients**

		Chi-square	df	Sig.
Step 1	Step	29.206	4	.000
	Block	29.206	4	.000
	Model	72.155	13	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	476.083 ^a	.166	.222

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10.043	8	.262

Contingency Table for Hosmer and Lemeshow Test

		Q88:Helped circulate,signed petition =		Q88:Helped circulate,signed petition =		Total
		no		yes		
		Observed	Expected	Observed	Expected	
Step 1	1	31	30.605	9	9.395	40
	2	31	27.469	9	12.531	40
	3	27	25.123	13	14.877	40
	4	21	22.878	19	17.122	40
	5	16	20.309	24	19.691	40
	6	12	18.167	28	21.833	40
	7	17	15.758	23	24.242	40
	8	16	12.143	24	27.857	40
	9	9	8.085	31	31.915	40
	10	4	3.463	33	33.537	37

Classification Table^a

Observed			Predicted		Percentage Correct
			Q88:Helped circulate,signed petition		
			no	yes	
Step 1	Q88:Helped circulate,signed petition	no	123	61	66.8
		yes	65	148	69.5
Overall Percentage					68.3

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	q104	-.021	.073	.079	1	.778	.980	.849	1.131
	q105	.146	.093	2.445	1	.118	1.157	.964	1.390
	whiteness	-.181	.271	.444	1	.505	.835	.491	1.420
	femaleness	-.040	.228	.030	1	.862	.961	.615	1.503
	q24	.033	.036	.796	1	.372	1.033	.962	1.110
	q52	.507	.247	4.207	1	.040	1.660	1.023	2.696
	q83	.324	.258	1.580	1	.209	1.383	.834	2.292
	q23	-.038	.041	.877	1	.349	.963	.889	1.043
	q32	-.015	.038	.163	1	.686	.985	.915	1.060
	q46	1.127	.398	8.016	1	.005	3.085	1.414	6.729
	q85	.625	.246	6.458	1	.011	1.868	1.154	3.025
	q86	.392	.334	1.378	1	.240	1.481	.769	2.851
	q90	.426	.288	2.195	1	.138	1.531	.872	2.690
	Constant	-1.189	.635	3.504	1	.061	.305		

a. Variable(s) entered on step 1: q46, q85, q86, q90.

IV. Tabling Results

Table 1: Prediction of signing or circulating a petition via Logistic Regression

					Model Statistics				
	r	Exp (B) in	Final Exp (B)	Block Chi-Sq	Model Chi-Sq	Model -2LL	Cox & Snell R ²	Nag. R ²	Hosmer & Lemeshow Chi-Sq
Block1: Demographics				11.23*	11.23*	537.01	.03	.04	9.52
Education	.14**	1.25**	1.16						
Whiteness	-.04	.81	.84						
Age	-.03	.98	.98						
Femaleness	-.08	.76	.96						
Block2: Community Participation				29.48**	40.71**	507.53	.10	.13	5.76
Q24	.16**	1.08*	1.03						
Q52	.25**	2.24**	1.66*						
Q83	.20**	1.43	1.38						
Block3: Sense of Gov. Security				2.24	42.95**	505.29	.10	.14	12.59
Q23	-.00	.95	.96						
Q32	-.11*	.97	.99						
Block4: Political Involvement				29.21**	72.16**	476.08	.17	.22	10.04
Q46	.26**	3.09**	3.09**						
Q85	.24**	1.87*	1.87*						
Q86	.19**	1.48	1.48						
Q90	.25**	1.53	1.53						

*p < .05; **p < .01

Table 2: Classification Results

			Final Predicted Group (Blocks 1 -4)		
			Q88: Helped circulate or signed a petition		Percentage Correct
			No	Yes	
Q88: Helped circulate or signed a petition	No	123	61	66.8%	
	Yes	65	148	69.5%	
Overall Percentage				68.3%	

Press' Q Calculation

Formula: $[N-(nK)]^2 / N(K-1)$

Where N= total sample size

n=number of observations correctly classified

K=number of groups

In our model:

$$N = 397$$

$$n = 271 [123+148]$$

$$K = 2$$

$$[397-(271*2)]^2 / 397(2-1)$$

$$21025/397$$

$$\text{Press' Q} = 52.96$$

$$df = 1$$

$$\text{Chi-sqcrit} = 10.83; p = .001$$

V. Write Up of Results

We used logistic regression to predict the likelihood of someone circulating or signing a petition, based on variables concerning community participation, sense of government security, and political involvement, as well as education, ethnicity, age, and gender. We decided to use hierarchical regression, which led to us placing our independent variables into four blocks. Block 1 was titled demographics and included variables related to demographics including education level, race (dummy coded as white = 1; all others = 0), age, and gender (dummy coded as femaleness = 1; male = 0). Block 2 was titled community participation because it contained variables associated with individuals participating in their community such as “I generally discuss political candidates and issues with neighbors at election time, have you worked with others in your community to solve some community problem in the past couple years, and have you attended meetings of your town or city council. Block three was titled sense of government security because it contained variables pertaining to individual’s faith in the government, such as public officials in my community seem receptive to views of residents, and other than voting, people like me have little influence over local government actions. Finally, Block four was titled political involvement because it contained variables related to individuals being engaged and involved in politics, such as being involved in political clubs or organizations, wearing a button or putting a sticker on your car, voting in the 2004 Presidential election, and contributing money to a party or candidate.

Table 1 summarizes our findings. Three of our blocks had significant Chi-squares: Block 1 (demographics) $\text{chi-sq.} = 11.23, p < .05$, Block 2 (community participation), $\text{chi-sq.} = 29.49, p < .01$ and Block 4 (political involvement) $\text{chi-sq.} = 29.21, p < .01$. Each Exp(B) indicates an increase or decrease in the odds of the occurrence of the dependent variable, assuming all other

independent variables in the model at that point are controlled for. Within our three significant blocks, three independent variables have significant unique contributions to the prediction of circulating or signing a petition. All three of these Exp(Bs) are positive, which indicates that they increase the odds of the dependent variable occurring, when controlling for all other independent variables in the model. The variable Q52 “worked with others on community problems” had a final Exp(B) of 1.66, which means for every increase in one unit of working with others on community problems, there will be a predicted increase of 66% in the odds of circulating or signing a petition, when all other independent variables are controlled for. The variable Q46 “belong to a political club or organization” had a final Exp(B) of 3.09, which means for every increase in 1 unit of belonging to a political club or organization, there will be a predicted increase of 209% in the odds of circulating or signing a petition, when all other independent variables are controlled for. Finally, the variable Q86 “voted in 2004 presidential election” had a final Exp(B) of 1.48, which means for every increase in 1 unit of voting in the 2004 presidential election, there will be a predicted increase of 48% in the odds of circulating or signing a petition, when all other independent variables are controlled for. Table 1 also reveals that all of our Hosmer and Lemeshow Chi-squares are non-significant.

Table 2 shows that our model correctly classified 68.3% of the cases. This is supported by the Press’ Q and Critical Chi-square statistics. Press’ Q was calculated at 52.96, while the Critical Chi-square for $p = .001$ is 10.83. Thus, since our calculated for Press’ Q exceeds the Critical Chi-square, the accuracy of our predictions is significantly greater than what could be expected by chance.

Overall, our four block model significantly predicts the likelihood of circulating or signing a petition. The full model -2LL is 476.08, with a Cox and Snell R^2 of .17 and a

Nagelkerke R^2 of .22. Also, our model suggests that when all other independent variables are controlled, for individuals who work with others on community problems, belong to a political club or organization, and voted in the 2004 Presidential election are more likely to circulate or sign a petition.