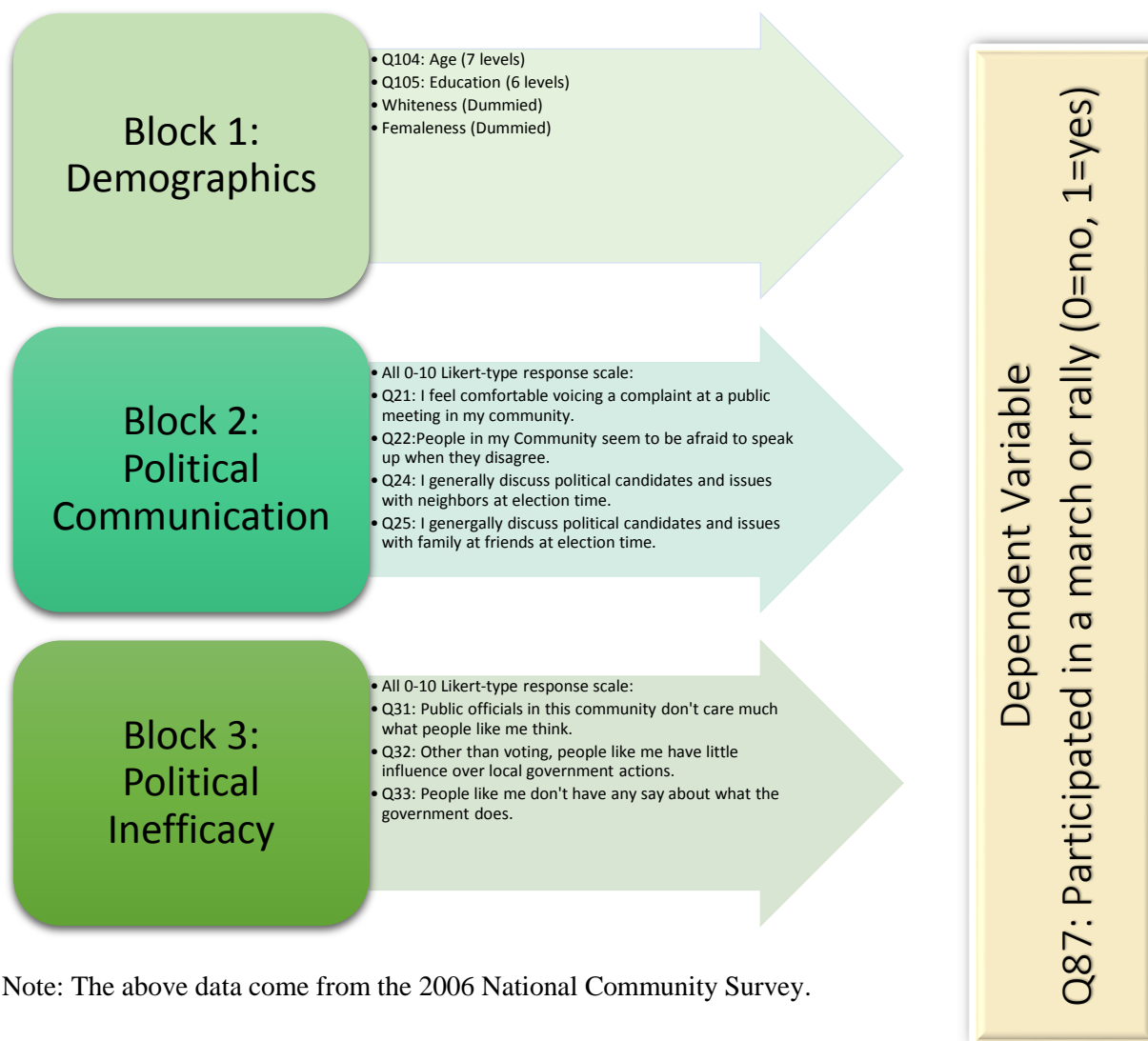


Logistic Regression

I. The Model

Independent Variables



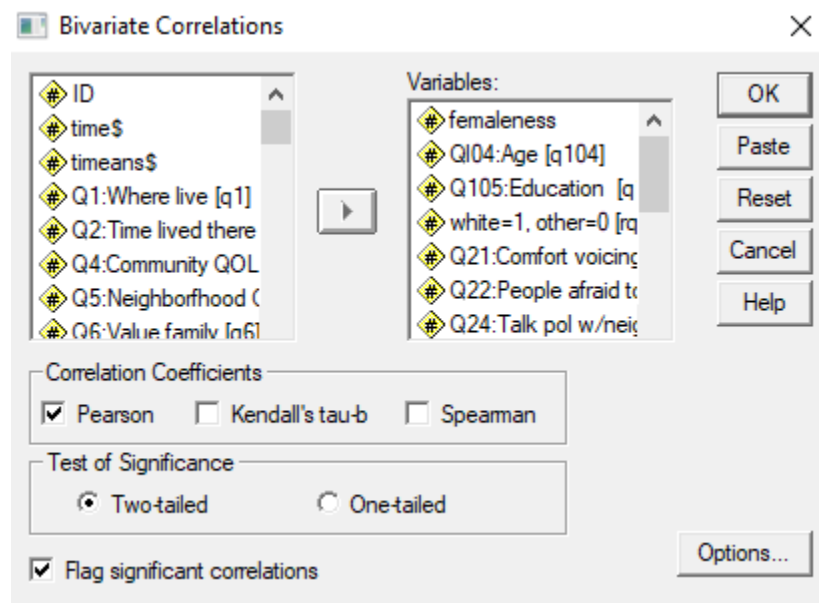
Note: The above data come from the 2006 National Community Survey.

II. Running SPSS (Logistic Regression)

First, run a basic Pearson's r correlation to look at correlations between each independent variable and the dependent variable as shown below.

Analyze → Correlate → Bivariate

Enter all of your variables in together.



Why do you think we do this?

Onto the fun part, Logistic Regression

Next, run a logistic regression for each block:

Analyze → Regression → Binary Logistic

- Select the dependent variable (Q87), then place each independent variable from Block 1 into the “covariates” section.
- Repeat for each block. (Each time this is done, SPSS will automatically view the covariates entered as one block.)
- SPSS also assumes a hierarchical ordering of the blocks, meaning each set of covariates entered as a block will be regressed to the dependent variable in the order the blocks are created.

For each block of variables select the dependent variable and the block's independent variables and press next:

Block One: Demographics

Logistic Regression

Dependent: Q87:Participated in ma

Block 1 of 3

Covariates: q104, q105, femaleness, rq106

Method: Enter

Selection Variable: Rule...

Categorical... Save... Options...

Additionally, within each block there is the option for stepwise or forced entry.

- Stepwise instructs SPSS to carry forward into the regression equation only the independent variable(s) which were found to be significant in relation to the dependent variable. This was what Dr. Neuendorf meant in the example of throwing things at the wall and seeing what sticks.
- We chose forced entry for all blocks, which instructs SPSS to carry all of the blocks' variables into the regression equation regardless of whether or not each individual variable is found to be significant. This is would be like nailing the variables to the wall.



Do you think this choice played a role in the significance of our variables or findings?

Block Two: Communication Likelihood

Logistic Regression

Dependent: Q87:Participated in ma

Block 2 of 3

Covariates: q21, q22, q24, q25

Method: Enter

Selection Variable: Rule...

Categorical... Save... Options...

Block Three: Political Inefficacy

Logistic Regression

Dependent:

Block 3 of 3

Covariates: q31, q32, q33

Method: Enter

Selection Variable: Rule...

Categorical... Save... Options...

→ “Options” that should be selected:

Logistic Regression: Options

Statistics and Plots

Classification plots Correlations of estimates

Hosmer-Lemeshow goodness-of-fit Iteration history

Casewise listing of residuals CI for exp(B): 95 %

All cases

Outliers outside 2 std. dev.

Display

At each step At last step

Probability for Stepwise

Entry: .05 Removal: .10

Classification cutoff: .5

Maximum iterations: 20

Include constant in model

Continue Cancel Help

→ NOW Click “Continue,” then click “Paste” on the main window, and run via syntax.

III. SPSS Output

```

q109
(2=1) (1=0) INTO Femaleness .
EXECUTE .
CORRELATIONS
/VARIABLES=q87 q104 q105 rq106 Femaleness q21 q22 q24 q25 q31 q32 q33
/PRINT=TWOTAIL NOSIG
/MISSING=PAIRWISE .

```

Correlations

Notes		
Output Created		08-APR-2018 13:20:56
Comments		
Input	Data	C:\Users\matta\Downloads\natcom (2).sav
	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.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
Syntax	CORRELATIONS /VARIABLES=q87 q104 q105 rq106 Femaleness q21 q22 q24 q25 q31 q32 q33 /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE .	
Resources	Elapsed Time	0:00:00.02

Correlations													
		Q87:Participated in march, rally	Q104:Age	Q105:Education	white=1, other=0	Femaleness	Q21:Comfort voicing complaints public meeting	Q22:People afraid to speak up	Q24:Talk pol w/neighbors election time	Q25:Talk pol w/family, friends election time	Q31:Public officials don't care what I think	Q32:Have little influence over local gov	Q33:Don't have say about what gov does
Q87:Participated in march, rally	Pears on Correlation	1	-.149(**)	.053	-.081	-.081	.104(*)	-.052	.088	.076	-.087	-.104(*)	-.103(*)
	Sig. (2-tailed)	.	.002	.271	.093	.089	.031	.282	.067	.112	.070	.031	.032
	N	442	431	428	426	436	430	423	440	440	432	435	438
Q104:Age	Pears on Correlation	-.149(**)	1	.018	.265(**)	.034	.155(**)	.078	.121(*)	.045	.026	.122(*)	.086
	Sig. (2-tailed)	.002	.	.712	.000	.483	.001	.114	.012	.347	.588	.012	.075
	N	431	433	427	424	431	421	414	431	431	424	427	429
Q105:Education	Pears on Correlation	.053	.018	1	.085	-.066	.273(**)	-.136(*)	.107(*)	.207(**)	-.177(*)	-.224(**)	-.233(**)
	Sig. (2-tailed)	.271	.712	.	.079	.173	.000	.006	.027	.000	.000	.000	.000
	N	428	427	429	422	427	418	410	427	427	420	423	425
white=1, other=0	Pears on Correlation	-.081	.265(**)	.085	1	.031	.057	-.060	.004	.036	-.010	-.041	-.002
	Sig. (2-tailed)	.093	.000	.079	.	.526	.247	.228	.941	.455	.845	.398	.973
	N	426	424	422	427	425	417	410	425	425	418	422	423
Femaleness	Pears on Correlation	-.081	.034	-.066	.031	1	.048	.080	.141(*)	-.052	.099(*)	-.012	-.004

	Sig. (2-tailed)	.089	.483	.173	.526	.	.324	.099	.003	.276	.039	.800	.932
	N	436	431	427	425	441	429	422	439	439	430	434	437
Q21: Comfort voicing complaints public meeting	Pears on Correlation	.104(*)	.155(**)	.273(**)	.057	.048	1	-.013	.291(*)	.321(**)	-.086	-.088	-.090
	Sig. (2-tailed)	.031	.001	.000	.247	.324	.	.782	.000	.000	.072	.065	.059
	N	430	421	418	417	429	449	433	447	446	437	441	443
Q22: People afraid to speak up	Pears on Correlation	-.052	.078	-.136(**)	-.060	.080	-.013	1	.062	-.001	.258(*)	.281(**)	.158(**)
	Sig. (2-tailed)	.282	.114	.006	.228	.099	.782	.	.191	.975	.000	.000	.001
	N	423	414	410	410	422	433	443	442	440	433	437	438
Q24: Talk pol w/neighbors election time	Pears on Correlation	.088	.121(*)	.107(*)	.004	.141(*)	.291(**)	.062	1	.491(**)	.043	-.040	-.026
	Sig. (2-tailed)	.067	.012	.027	.941	.003	.000	.191	.	.000	.359	.394	.584
	N	440	431	427	425	439	447	442	459	457	448	451	454
Q25: Talk pol w/family, friends election time	Pears on Correlation	.076	.045	.207(**)	.036	-.052	.321(**)	-.001	.491(*)	1	-.068	-.089	-.128(**)
	Sig. (2-tailed)	.112	.347	.000	.455	.276	.000	.975	.000	.	.152	.060	.006
	N	440	431	427	425	439	446	440	457	459	448	451	455
Q31: Public officials don't care what I think	Pears on Correlation	-.087	.026	-.177(**)	-.010	-.099(*)	-.086	.258(*)	.043	-.068	1	.489(**)	.454(**)
	Sig. (2-tailed)	.070	.588	.000	.845	.039	.072	.000	.359	.152	.	.000	.000
	N	432	424	420	418	430	437	433	448	448	449	445	447
Q32: Have little influence	Pears on Correlation	-.104(*)	.122(*)	-.224(**)	-.041	-.012	-.088	.281(*)	-.040	-.089	.489(*)	1	.557(**)

over local gov	Sig. (2-tailed)	.031	.012	.000	.398	.800	.065	.000	.394	.060	.000	.	.000
	N	435	427	423	422	434	441	437	451	451	445	453	451
Q33:Don't have say about what gov does	Pears on Correlation	-.103(*)	.086	-.233(**)	.002	-.004	-.090	.158(*)	-.026	.128(**)	.454(*)	.557(**)	1
	Sig. (2-tailed)	.032	.075	.000	.973	.932	.059	.001	.584	.006	.000	.000	.
	N	438	429	425	423	437	443	438	454	455	447	451	456
** Correlation is significant at the 0.01 level (2-tailed).													
* Correlation is significant at the 0.05 level (2-tailed).													

```
LOGISTIC REGRESSION q87
  /METHOD = ENTER q104 q105 rq106 Femaleness /METHOD = ENTER q21 q22 q24 q25
/
  METHOD = ENTER q31 q32 q33
/CLASSPLOT
/PRINT = GOODFIT CI(95)
/CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
```

Logistic Regression

Notes		
Output Created		08-APR-2018 13:26:38
Comments		
Input	Data	C:\Users\matta\Downloads\natcom (2).sav
	File Label	CP05
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
Syntax		LOGISTIC REGRESSION q87 /METHOD = ENTER q104 q105 rq106 Femaleness /METHOD = ENTER q21 q22 q24 q25 /METHOD = ENTER q31 q32 q33 /CLASSPLOT /PRINT = GOODFIT CI(95) /CRITERIA = PIN(.05) POUT(.10) ITERATE(20) CUT(.5) .
Resources	Elapsed Time	0:00:00.06

Case Processing Summary			
Unweighted Cases(a)		N	Percent
Selected Cases	Included in Analysis	385	80.7
	Missing Cases	92	19.3
	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
0=no	0
1=yes	1

Block 0: Beginning Block

Classification Table(a,b)					
	Observed		Predicted		Percentage Correct
			Q87:Participated in march, rally		
			0=no	1=yes	
Step 0	Q87:Participated in march, rally	0=no	332	0	100.0
		1=yes	53	0	.0
	Overall Percentage				

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	-1.835	.148	153.869	1	.000	.160

Variables not in the Equation					
		Score	df	Sig.	
Step 0	Variables	q104	5.155	1	.023
		q105	3.789	1	.052
		rq106	.263	1	.608

		Femaleness	4.791	1	.029
	Overall Statistics		13.142	4	.011

Block 1: Method = Enter

		Chi-square	df	Sig.
Step 1	Step	14.382	4	.006
	Block	14.382	4	.006
	Model	14.382	4	.006

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	294.155(a)	.037	.067

a Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Step	Chi-square	df	Sig.
1	4.752	8	.784

		Q87:Participated in march, rally = 0=no		Q87:Participated in march, rally = 1=yes		Total
		Observed	Expected	Observed	Expected	
Step 1	1	37	37.423	2	1.577	39
	2	41	39.007	1	2.993	42
	3	35	35.551	4	3.449	39
	4	36	36.785	5	4.215	41
	5	34	33.268	4	4.732	38
	6	35	34.195	5	5.805	40
	7	29	32.518	10	6.482	39
	8	33	31.841	6	7.159	39
	9	30	30.334	9	8.666	39
	10	22	21.076	7	7.924	29

Each Symbol Represents 5 Cases.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	9.992	4	.041
	Block	9.992	4	.041
	Model	24.374	8	.002

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	284.162(a)	.061	.111

a Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	20.846	8	.008

Contingency Table for Hosmer and Lemeshow Test						
		Q87:Participated in march, rally = 0=no		Q87:Participated in march, rally = 1=yes		Total
		Observed	Expected	Observed	Expected	
Step 1	1	37	37.948	2	1.052	39
	2	38	37.009	1	1.991	39
	3	38	36.228	1	2.772	39
	4	39	35.548	0	3.452	39
	5	35	34.654	4	4.346	39
	6	26	33.837	13	5.163	39
	7	32	32.832	7	6.168	39
	8	33	31.530	6	7.470	39
	9	30	29.577	9	9.423	39
	10	24	22.837	10	11.163	34

Classification Table(a)					
	Observed		Predicted		Percentage Correct
			Q87:Participated in march, rally		
			0=no	1=yes	
Step 1	Q87:Participated in march, rally	0=no	332	0	100.0
		1=yes	53	0	.0
	Overall Percentage				86.2

a The cut value is .500

Block 3: Method = Enter

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	2.945	3	.400
	Block	2.945	3	.400
	Model	27.320	11	.004

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	281.217(a)	.069	.124

a Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.203	8	.625

Contingency Table for Hosmer and Lemeshow Test						
		Q87:Participated in march, rally = 0=no		Q87:Participated in march, rally = 1=yes		Total
		Observed	Expected	Observed	Expected	
Step 1	1	37	37.992	2	1.008	39
	2	38	37.164	1	1.836	39
	3	38	36.536	1	2.464	39
	4	36	35.670	3	3.330	39
	5	35	34.799	4	4.201	39
	6	34	33.893	5	5.107	39
	7	32	32.876	7	6.124	39
	8	28	31.275	11	7.725	39
	9	28	29.379	11	9.621	39
	10	26	22.416	8	11.584	34

Classification Table(a)					
	Observed		Predicted		Percentage Correct
			Q87:Participated in march, rally		
			0=no	1=yes	
Step 1	Q87:Participated in march, rally	0=no	330	2	99.4
		1=yes	53	0	.0
	Overall Percentage				85.7

a The cut value is .500

IV. Tabling Results

Table 1: Prediction of Self-Reported Participation in a March or Rally via Logistic Regression

	r	Final Exp (B)	Block Chi-Sq	Model Chi-Sq	Model -2LL	Cox & Snell R ²	Nag R ²	Hosmer & Lemeshow Chi-Sq
Block 1: Demographics			14.382**	14.382**	294.155	0.037	0.067	4.751
Q104. Age	-0.149**	0.753**						
Q105. Education completed	0.053	1.114						
RQ106. Race (white=1, other=0)	-0.081 ^a	0.887						
RQ109. Femaleness	-0.081 ^a	0.515*						
Block 2: Political Communication			9.992*	24.374**	284.162	0.061	0.111	20.845**
Q21: I'd feel comfortable voicing a complaint at a public meeting in my community.	0.104*	1.110 ^a						
Q22: People in this community seem to be afraid to speak up when they disagree.	-0.052	0.973						
Q24. I generally discuss political candidates and issues with neighbors at election time.	0.088 ^a	1.066						
Q25. I generally discuss political candidates and issues with family and friends at election time.	0.076	1.016						
Block 3: Political Inefficacy			2.945	27.320**	281.217	0.069	0.124	6.203
Q31: Public officials don't care much what people like me think.	-0.087 ^a	0.996						
Q32: Other than voting, people like me have little influence over local government actions.	-0.104*	0.946						
Q33: People like me don't have any say about what the government does.	-0.103*	0.960						

*** Correlation is significant at $p < 0.001$ level (2-tailed)

** Correlation is significant at $p \leq 0.01$ level (2-tailed)

* Correlation is significant at $p \leq 0.05$ level (2-tailed)

^a $.05 < p < .10$

Table 2: Classification Results(a)

Observed		Predicted		
		Q87: Participated in march, rally		Percentage Correct
		No= 383	Yes= 2	
Q87: Participated in march, rally	No= 332	330	2	99.4
	Yes= 53	53	0	0.0
Overall Percentage				85.7

a The cut value is .500

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 this model:

N=385

n = 330 + 0 = 330

K = 2

$$\begin{aligned} \text{Press' Q} &= [385-(330*2)]^2 / 385(2-1) \\ &= [385-660]^2 / 385 \\ &= 75,625 / 385 \end{aligned}$$

$$\text{Press' Q} = 196.4 \quad \text{df} = 1$$

Critical chi-square at 0.001 level of significance = 10.83

V. Results Write-Up

To predict the likelihood of someone participating in a march or rally given a chosen set of variables, we used logistic regression. All data came from the 2006 National Community Study class data set. We grouped the independent variables into blocks so that the model could be run hierarchically. Block 1 contained the personal demographic variables and thus we named this block “demographics” to characterize the block’s variables, which describe a person’s age, education, race and gender. Both race and gender were dummy coded to represent whiteness and femaleness, respectively. Block 2 was titled “Political Communication.” This contained variables related to political discussion and perceived climate for communication in one’s community. Block 3 was developed based on the 2006 National Community Study scale for “Perceived Political Efficacy,” however since all three items are negatively worded we changed the name of this block to “Political Inefficacy.” We chose to use the forced entry method for each of these blocks in the logistic regression. Forced entry instructs SPSS to use all variables in the block regardless of the significance of each individual variable.

As indicated in Table 1, Q104: Age had the most significant correlation (r) to Q87: Participated in a march or rally, at $r = -.149$, $p < 0.01$ level. Three independent variables were significant at $p < .05$: Q21: I’d feel comfortable voicing a complaint at a public meeting in my community ($r = .104$); Q32: Other than voting, people like me have little influence over local government actions ($r = -.104$); Q33: People like me don’t have any say about what the government does ($r = -.103$). Other variables with near significance at the $0.05 < p < 0.10$ level were: RQ106: Recoded Race (whiteness) ($r = -.081$); RQ109: Femaleness ($r = -.081$); and Q31: Public officials don’t care much what people like me think ($r = -.087$).

Block 1 contributed to the prediction of participation in a march or rally significantly, with a Chi-square for the block of 14.382 ($p < .01$). In Block 1, only Q104: Age and RQ109: Femaleness, had significant final Exp(B)s (0.753 and 0.515), which indicated a 25% decrease in the odds of a person participating in a march or rally for each unit increase in age and a 48% lower odds of participating for female respondents (vs. males), when all other independent variables were controlled for. (Note: Q104 was measured on a 7-point response scale.)

Block 2 was found to have a significant block Chi-square of 9.992 ($p < .05$). As the model was run hierarchically, the addition of Block 2 increased the model Chi-square to 24.374, which was also significant ($p < .01$). We used the forced entry method, so all variables were included in the equation but only one of the four had a near significant final Exp(B). The most significant final Exp(B) in Block 2 was the one that also had a significant correlation; the final Exp(B) of 1.110 indicated a 11% increase in the odds someone will participate in a march or rally for each unit increase in the measure of how much more they tend to feel comfortable voicing a complaint at a public meeting in their community (when all other independent variables were controlled for; Q21 was measured on a 0-to-10 response scale).

Moving to Block 3 we begin to see how a hierarchical model may impact the big picture. All three of the variables in Block 3 reflected a moderately to slightly significant r , i.e., Q31, Q32, and Q33 all were significant at the $p < .05$ level or near-significant at $.05 < p < .10$. One might assume this block to have at least a slightly significant impact on the overall model. However, as we see in Table 1, the Block 3 overall Chi-square of 2.945 was not significant. The model Chi-square remained significant, and did increase to 27.320 ($p < .01$), but perhaps not as much of an increase as we may have predicted. None of the variables in Block 3 had a significant final Exp(B). We might suspect this is due to the hierarchical nature of the model, which would not allow for a strong regression of the Block 3 variables if those variables have a great deal of “overlap” with the Block 1 or Block 2 variables. If the model were to be run by switching the order of Blocks 2 and 3, perhaps we would find that block to be significant.

Table 1 also reveals that the Hosmer & Lemeshow goodness-of-fit test (another assessment of how well the model fits the data) was found to be non-significant at Blocks 1 and 3, but was significant at $p < 0.01$ level for Block 2, which indicates not a good fit for the model overall with two blocks in. The -2LL for the full model is 281.217, which, given its high dependence on n , is often thought to be better interpreted by Cox & Snell R^2 and Nagelkerke R^2 . The Cox & Snell R^2 value of 0.069 with all three blocks in indicated the independent variables in the full model explained approximately 7% of the variance in the dependent variable. This is further confirmed by the Nagelkerke R^2 of 0.124 for the full model, estimating 12% of the variance of the dependent variable was explained by the independent variables included in the overall model.

As shown in Table 2, the model correctly classified 85.7% of the cases. The Press' Q calculation of 196.4 supports this finding, as it exceeds the critical chi-square of 10.83 at the 0.001 significance level. Therefore, the accuracy of the model's predictions is significantly greater than what could be expected by chance.