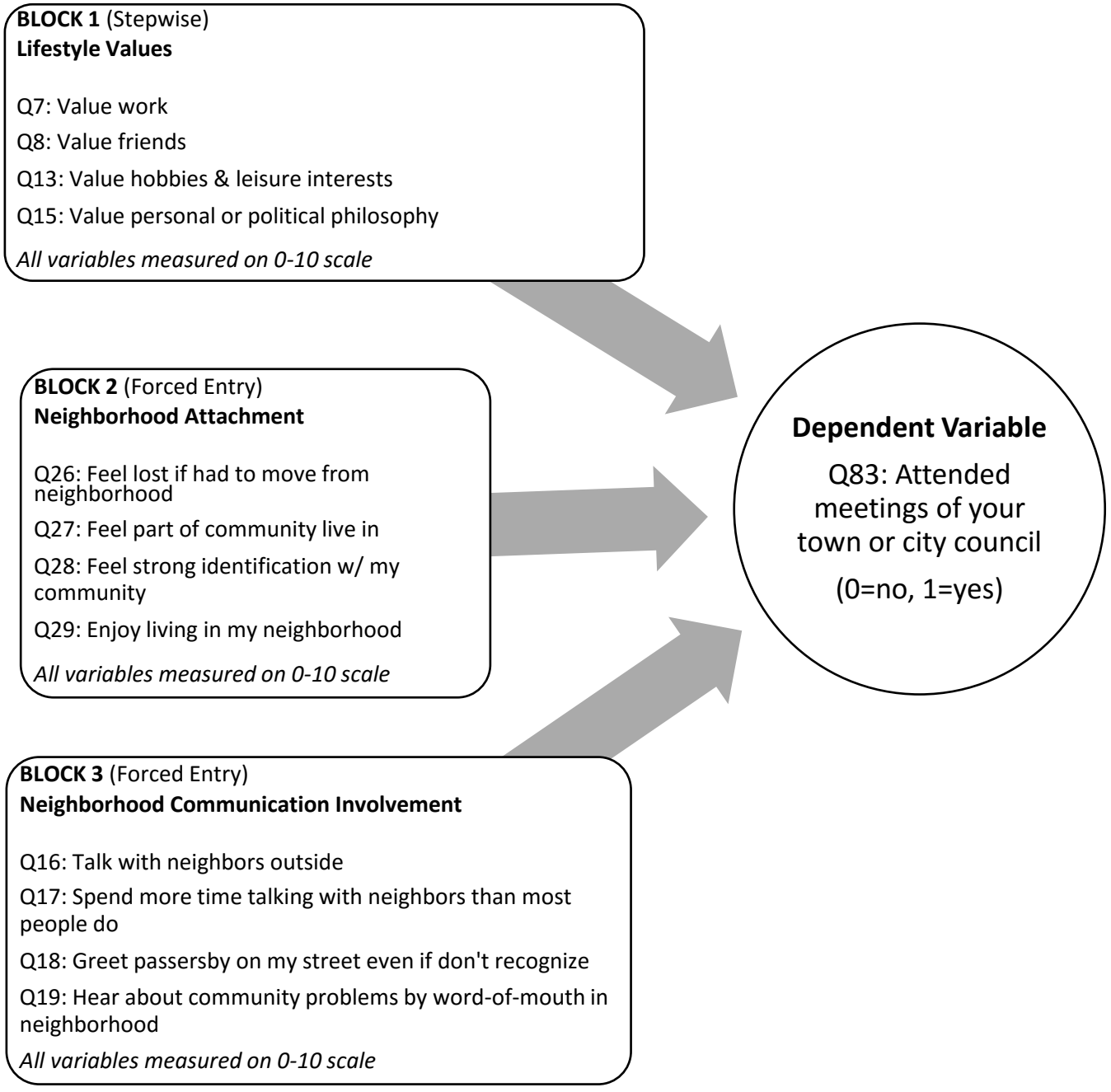
Kristi Andrasik
COM 731
Spring 2017

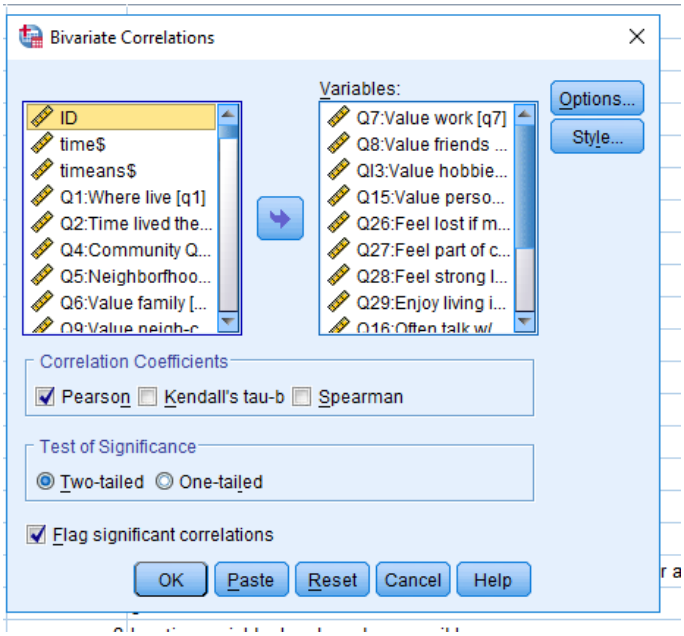
Logistic Regression

I. MODEL – all data drawn from the 2006 National Community Survey (class data set)



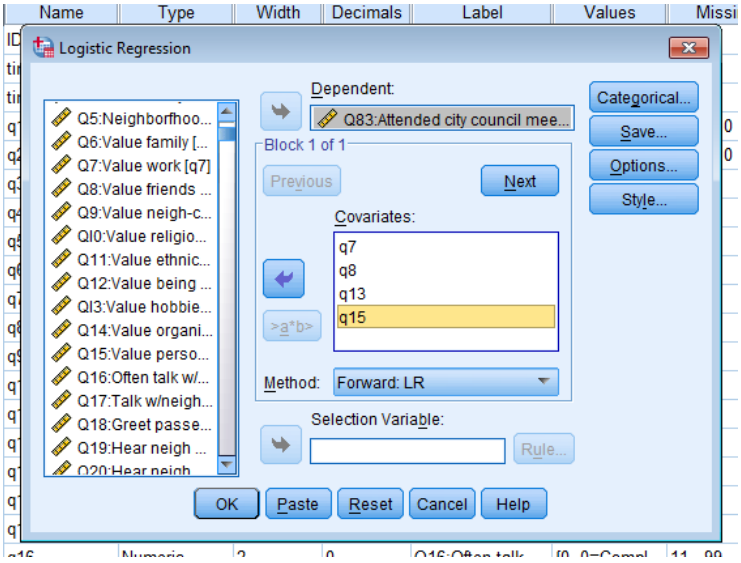
II. Running SPSS

First, run a basic Pearson's r correlation to look at correlations between each independent variable and the dependent variable. Analyze → Correlate → Bivariate

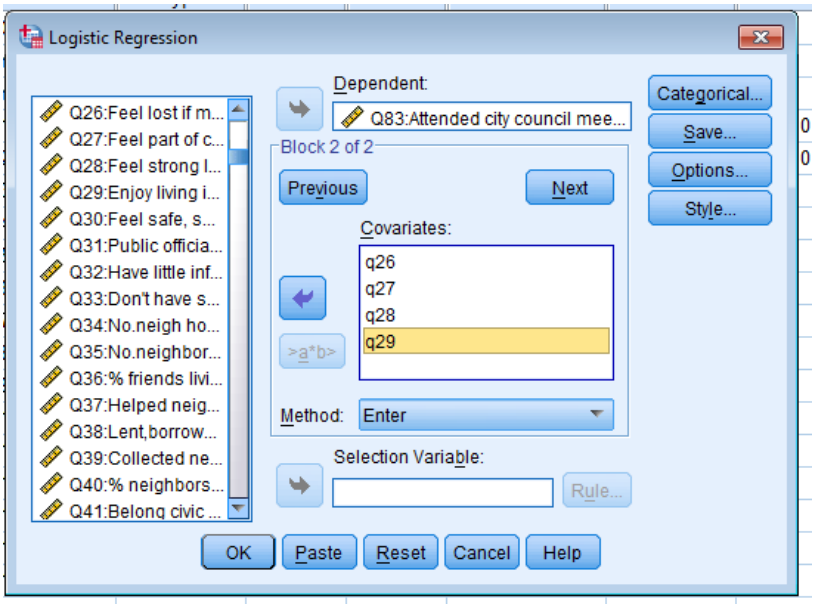


Next, run a logistic regression for each block via Analyze → Regression → Binary Logistic. Select the dependent variable (Q83), then place each independent variable from Block 1 into the "covariates" section. Repeat for each block. Each time this is done, SPSS will automatically view the covariates entered 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. Additionally, within each block there is the option for stepwise or forced entry. In my first block, I chose stepwise ("Forward: LR"). This 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. I then chose forced entry for my second and third blocks, which instructs SPSS to carry all of those blocks' variables into the regression equation regardless of whether or not each individual variable is found to be significant.

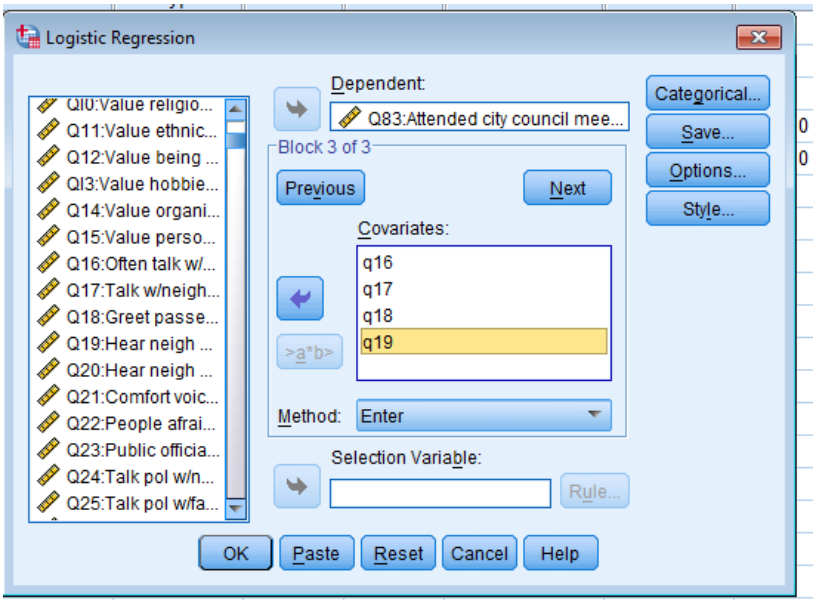
Block 1:



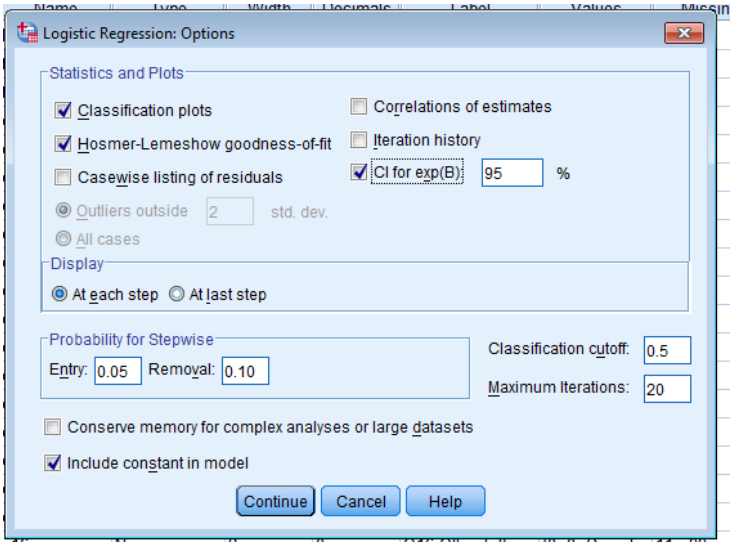
Block 2:



Block 3:



For each block, ensure the following options are selected:



III. SPSS Output

Note: for the sake of space, I have included the correlation output only between the dependent variable and each of the independent variables in my model. I did not include the correlations of independent variable-to-independent variable.

Correlations

Notes

Output Created		10-APR-2017 20:56:56
Comments		
Input	Data	C:\Users\2203198\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.
	Cases Used	Statistics for each pair of variables are based on all the cases with valid data for that pair.
Syntax		CORRELATIONS /VARIABLES=q7 q8 q13 q15 q26 q27 q28 q29 q16 q17 q18 q19 q83 /PRINT=TWOTAIL NOSIG /MISSING=PAIRWISE.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.01

		Q83:Attended city council meetings
Q7:Value work	Pearson Correlation	.004
	Sig. (2-tailed)	.932
	N	401
Q8:Value friends	Pearson Correlation	-.093
	Sig. (2-tailed)	.051
	N	438
Ql3:Value hobbies-leisure	Pearson Correlation	-.125**
	Sig. (2-tailed)	.009
	N	436
Q15:Value personal-pol.philosophy	Pearson Correlation	.021
	Sig. (2-tailed)	.670
	N	434
Q26:Feel lost if moved from neighborhood	Pearson Correlation	-.031
	Sig. (2-tailed)	.514
	N	438
Q27:Feel part of community	Pearson Correlation	.107*
	Sig. (2-tailed)	.025
	N	439
Q28:Feel strong ID w/community	Pearson Correlation	.078
	Sig. (2-tailed)	.104
	N	438
Q29:Enjoy living in neighborhood	Pearson Correlation	-.044
	Sig. (2-tailed)	.358
	N	439
Q16:Often talk w/neighbors on street	Pearson Correlation	.072
	Sig. (2-tailed)	.134
	N	435
Q17:Talk w/neighbors more than most	Pearson Correlation	.147**
	Sig. (2-tailed)	.002
	N	436
Q18:Greet passersby	Pearson Correlation	.099*
	Sig. (2-tailed)	.038
	N	439
Q19:Hear neigh problems word-of-mouth	Pearson Correlation	.094*
	Sig. (2-tailed)	.048
	N	439

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

```
LOGISTIC REGRESSION VARIABLES q83
  /METHOD=FSTEP(LR) q7 q8 q13 q15
  /METHOD=ENTER q26 q27 q28 q29
  /METHOD=ENTER q16 q17 q18 q19
  /CLASSPLOT
  /PRINT=GOODFIT CI(95)
  /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```

Logistic Regression

Notes

Output Created		05-APR-2017 19:05:05
Comments		
Input	Data	C:\Users\228 lab02\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.IE5\UEJ49OU8\natcom.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 q83 /METHOD=FSTEP(LR) q7 q8 q13 q15 /METHOD=ENTER q26 q27 q28 q29 /METHOD=ENTER q16 q17 q18 q19 /CLASSPLOT /PRINT=GOODFIT CI(95) /CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
Resources	Processor Time	00:00:00.06
	Elapsed Time	00:00:00.10

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	383	80.3
	Missing Cases	94	19.7
	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]

			Predicted		
			Q83:Attended city council meetings		Percentage
	Observed		0=no	1=yes	Correct
Step 0	Q83:Attended city council meetings	0=no	241	0	100.0
		1=yes	142	0	.0
	Overall Percentage				62.9

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	-.529	.106	25.002	1	.000	.589

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	q7	.032	1	.859
		q8	4.948	1	.026
		q13	6.582	1	.010
		q15	.022	1	.883
	Overall Statistics		9.529	4	.049

Block 1: Method = Forward Stepwise (Likelihood Ratio)

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	6.504	1	.011
	Block	6.504	1	.011
	Model	6.504	1	.011

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	498.564[a]	.017	.023

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10.111	5	.072

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

Contingency Table for Hosmer and Lemeshow Test

		Q83:Attended city council meetings = 0=no		Q83:Attended city council meetings = 1=yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	57	53.333	19	22.667	76
	2	25	27.766	16	13.234	41
	3	53	50.832	25	27.168	78
	4	34	43.767	36	26.233	70
	5	21	19.137	11	12.863	32
	6	37	31.932	19	24.068	56
	7	14	14.233	16	15.767	30

Classification Table[a]

			Predicted		
			Q83:Attended city council meetings		Percentage
	Observed		0=no	1=yes	Correct
Step 1	Q83:Attended city council meetings	0=no	235	6	97.5
		1=yes	130	12	8.5
	Overall Percentage				64.5

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
Step 1[a]	q13	-.115	.045	6.420	1	.011	.892	.816
	Constant	.290	.338	.736	1	.391	1.337	

Variables in the Equation

		95% C.I.for EXP(B) Upper
Step 1[a]	q13	.974
	Constant	

a. Variable(s) entered on step 1: q13.

Model if Term Removed

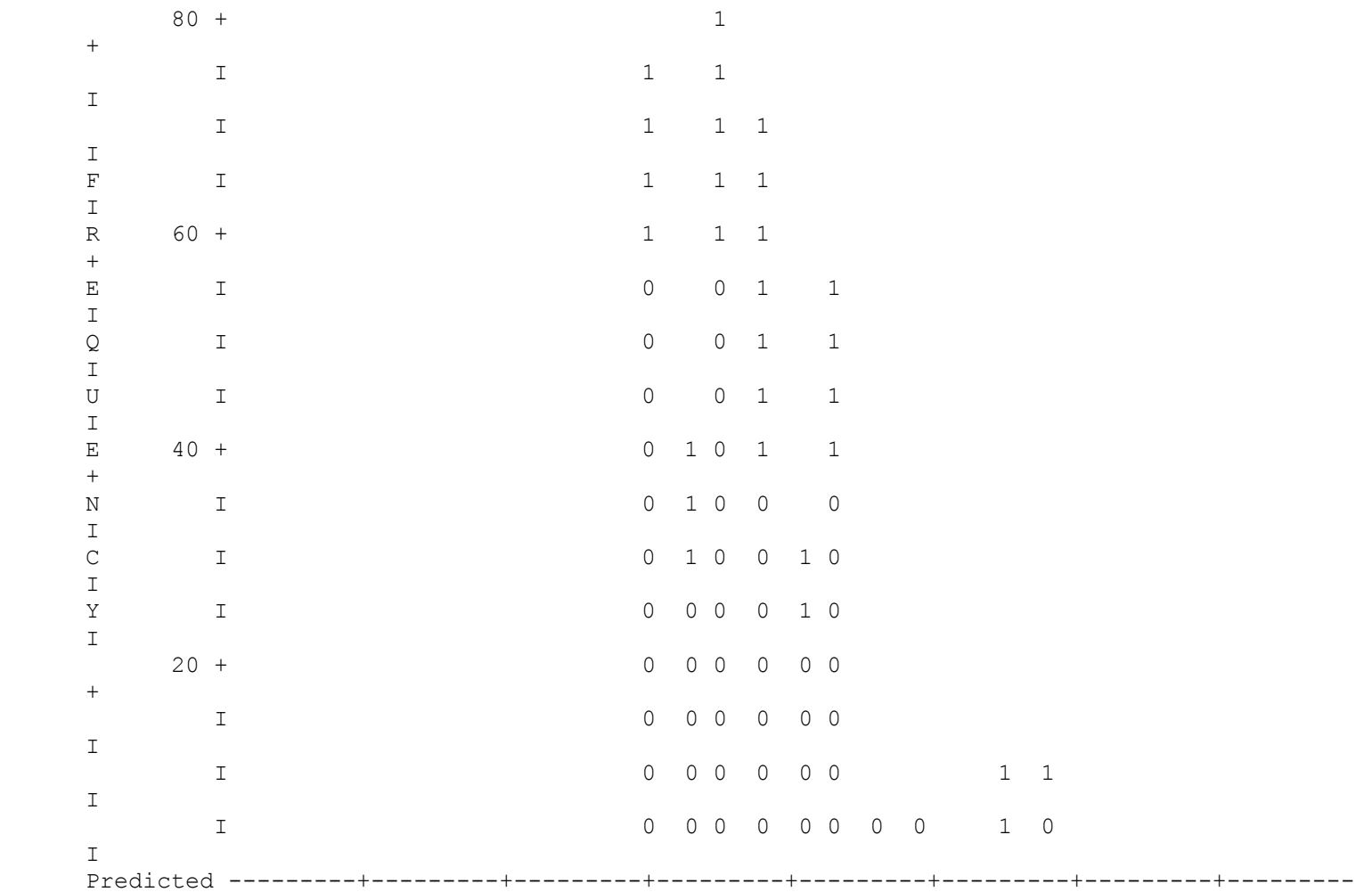
Variable		Model Log Likelihood	Change in -2 Log Likelihood	df	Sig. of the Change
Step 1	q13	-252.534	6.504	1	.011

Variables not in the Equation

			Score	df	Sig.
Step 1	Variables	q7	.031	1	.860
		q8	1.711	1	.191
		q15	.723	1	.395
	Overall Statistics		2.962	3	.397

Step number: 1

Observed Groups and Predicted Probabilities



Predicted Probability is of Membership for 1=yes
The Cut Value is .50
Symbols: 0 - 0=no
1 - 1=yes
Each Symbol Represents 5 Cases.

Block 2: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	24.561	4	.000
	Block	24.561	4	.000
	Model	31.065	5	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	474.003[a]	.078	.106

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	5.147	8	.742

Contingency Table for Hosmer and Lemeshow Test

		Q83:Attended city council meetings = 0=no		Q83:Attended city council meetings = 1=yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	34	32.962	4	5.038	38
	2	32	29.245	6	8.755	38
	3	28	27.388	10	10.612	38
	4	23	25.768	15	12.232	38
	5	22	24.390	16	13.610	38
	6	23	23.284	15	14.716	38
	7	21	23.360	19	16.640	40
	8	24	20.571	14	17.429	38
	9	20	18.793	18	19.207	38
	10	14	15.238	25	23.762	39

Classification Table[a]

Observed			Predicted		
			Q83:Attended city council meetings		Percentage Correct
			0=no	1=yes	
Step 1	Q83:Attended city council meetings	0=no	215	26	89.2
		1=yes	103	39	27.5
	Overall Percentage				66.3

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
Step 1[a]	q13	-.159	.052	9.454	1	.002	.853	.771
	q26	-.067	.034	3.926	1	.048	.935	.875
	q27	.233	.070	11.093	1	.001	1.263	1.101
	q28	.042	.062	.459	1	.498	1.043	.923
	q29	-.133	.069	3.693	1	.055	.876	.765
	Constant	.085	.436	.038	1	.845	1.089	

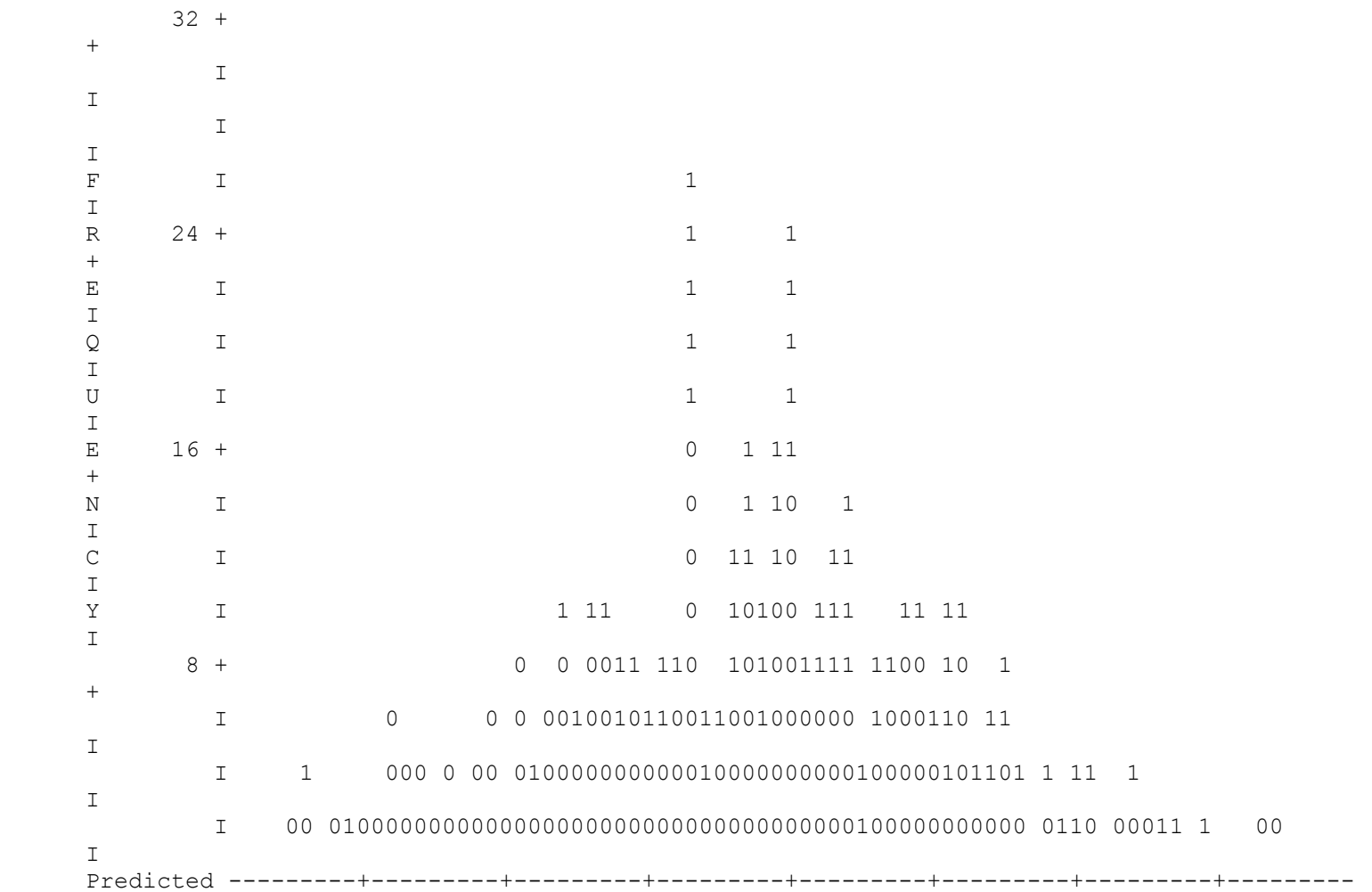
Variables in the Equation

		95% C.I.for EXP(B) Upper
Step 1[a]	q13	.944
	q26	.999
	q27	1.449
	q28	1.178
	q29	1.003
	Constant	

a. Variable(s) entered on step 1: q26, q27, q28, q29.

Step number: 1

Observed Groups and Predicted Probabilities



Predicted Probability is of Membership for 1=yes
The Cut Value is .50
Symbols: 0 - 0=no
1 - 1=yes
Each Symbol Represents 2 Cases.

Block 3: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.564	4	.109
	Block	7.564	4	.109
	Model	38.629	9	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	466.439[a]	.096	.131

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	10.816	8	.212

Contingency Table for Hosmer and Lemeshow Test

		Q83:Attended city council meetings = 0=no		Q83:Attended city council meetings = 1=yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	36	33.556	2	4.444	38
	2	31	29.766	7	8.234	38
	3	27	27.792	11	10.208	38
	4	26	26.525	12	11.475	38
	5	19	24.737	19	13.263	38
	6	28	23.218	10	14.782	38
	7	23	22.906	17	17.094	40
	8	17	20.557	21	17.443	38
	9	17	18.111	21	19.889	38
	10	17	13.833	22	25.167	39

Classification Table[a]

Observed			Predicted		
			Q83:Attended city council meetings		Percentage Correct
			0=no	1=yes	
Step 1	Q83:Attended city council meetings	0=no	209	32	86.7
		1=yes	102	40	28.2
	Overall Percentage				65.0

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
Step 1[a]	q13	-.170	.053	10.128	1	.001	.844	.760
	q26	-.064	.034	3.457	1	.063	.938	.877
	q27	.219	.071	9.515	1	.002	1.245	1.083
	q28	.012	.064	.033	1	.857	1.012	.892
	q29	-.140	.071	3.840	1	.050	.869	.756
	q16	-.052	.055	.913	1	.339	.949	.852
	q17	.078	.048	2.612	1	.106	1.081	.984
	q18	.070	.047	2.238	1	.135	1.073	.978
	q19	.037	.040	.825	1	.364	1.037	.959
	Constant	-.232	.462	.251	1	.616	.793	

Variables in the Equation

		95% C.I.for EXP(B) Upper
Step 1[a]	q13	.937
	q26	1.003
	q27	1.430
	q28	1.147
	q29	1.000
	q16	1.057
	q17	1.188
	q18	1.177
	q19	1.122
	Constant	

a. Variable(s) entered on step 1: q16, q17, q18, q19.

```
             Step number: 1

             Observed Groups and Predicted Probabilities

      20 +                                          1
+
         I                           1              1
I
         I                           1              1
I
F        I                           1              1
I
R     15 +                           0              1
+
E        I                           0 1          1 1
I
Q        I                           011      1   1 1 1
I
U        I                      1    011     10   1 111
I
E     10 +                      1    011    110   1 111
+
N        I                      0    011   1110   0 111
I
C        I                      0111 0000  110011101011 1  1 1
I
Y        I                      0110 0000  110010000011 1  1 1   11
I
       5 +           0       10  0100 0000  010010000000 11 1 01  11
+
         I           00 00 0 1001000010000110000000000000111 01011 11
I
         I      0000000000 000000000000000010000000000000000101010100001101     11 1
1                  I
         I      000000000000000000000000000000000000000000000000000001101  11 0011   0  0
01                 I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------
+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8
.9         1
  Group:
000000000000000000000000000000000000000000000000000011111111111111111111111111111111111111
1111111111

          Predicted Probability is of Membership for 1=yes
          The Cut Value is .50
          Symbols: 0 - 0=no
                   1 - 1=yes
          Each Symbol Represents 1.25 Cases.
```

IV. Tabling Results

Table 1: Prediction of attending a city council meeting 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: Lifestyle Values			6.50**	6.50**	498.56	0.02	0.02	10.11*
~~Q7: Value work~~	~~0.004~~							
~~Q8: Value friends~~	~~-0.093~~							
Q13: Value hobbies & leisure interests	-0.125 ***	0.84 ****						
~~Q15: Value personal or political philosophy~~	~~0.021~~							
Block 2: Neighborhood Attachment			24.56 ****	31.06 ****	474.00	0.08	0.106	5.15
Q26: Feel lost if had to move from neighborhood	-0.031	0.94*						
Q27: Feel part of community live in	0.107**	1.24 ***						
Q28: Feel strong identification w/ my community	0.078	1.01						
Q29: Enjoy living in my neighborhood	-0.044	0.87**						
Block 3: Neighborhood Communication Involvement			7.564	38.63 ****	466.44	0.10	0.13	10.82
Q16: Talk with neighbors outside	0.072	0.95						
Q17: Spend more time talking w/ neighbors than most people do	0.147 ***	1.08						
Q18: Greet passersby on my street even if don't recognize	0.099**	1.07						
Q19: Hear about community problems by word-of-mouth in neighborhood	0.094**	1.04						
	r	**Final Exp (B)**	**Block Chi-Sq**	**Model Chi-Sq**	**Model -2LL**	**Cox & Snell R^2**	**Nag R^2**	**Hosmer & Lemeshow Chi-Sq**

*p<0.10
**p≤0.05
***p<0.01
****p≤ 0.001

V. Results Write-Up

To predict the likelihood of someone attending a city council meeting given a chosen set of variables, I used logistic regression. All data came from the 2006 National Community Study class data set. I grouped the independent variables into blocks so that the model could be run hierarchically. Block 1 contained the variables chosen by a colleague who is also researching the factors that impact city council meeting attendance. I named this block "lifestyle values" to characterize the block's variables which each describe a component of social life that someone may value. I ran this block as a stepwise regression, which resulted in SPSS choosing just one of the variables in the block (Q13: value hobbies and leisure interests) to carry into the regression equation, as this was the only variable in Block 1 that was found to be significant. This finding is in alignment with my colleague's findings as well. Blocks 2 and 3 were developed based on the 2006 National Community Study scales for "Neighborhood Attachment" and "Involvement in Neighborhood Communication Network". Given the fact that these variables were already grouped into a scale by the National Community Study researchers, I 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. This led to some interesting findings.

As indicated in Table 1, Q13: valuing hobbies/leisure, was the only variable in Block 1 which had a significant correlation (r). This was also the only variable in Block 1 with a significant Exp(B), which indicates a 16% decrease in the odds of a person attending a city council meeting for each unit increase in the measure of how much they value hobbies/leisure (when all other independent variables are controlled for). (Note: Q13 was measured on an 11-point, 0-to-10 response scale.) Since this block was run as a stepwise regression, only this one significant variable was ultimately included in the block, resulting in a significant Chi-square for the block of 6.50.

Block 2 was found to have a highly significant Chi-square of 24.56. As the model is run hierarchically, the addition of Block 2 increased the model Chi-square to 31.06, which was also highly significant. Interestingly, just one of the four variables in Block 2 (Q27: feel part of community live in) was significantly correlated; however, since I used the forced entry method, all variables were included in the equation and three of the four have a significant Exp(B). The most significant Exp(B) in Block 2 is the one that also has a significant correlation; the Exp(B) indicates a 24% increase in the odds someone will attend a city council meeting for each unit increase in the measure of how much they feel part of the community they live in (when all other independent variables are controlled for; again, Q27 was measured on a 0-to-10 response scale). The other significant unique predictors in this block were Q26 (feel lost if had to move from neighborhood) and Q29 (enjoy living in my neighborhood). The Exp(B) of 0.94 for Q26 indicates a 6% decrease in the odds someone will attend a city council meeting for each unit increase in the measure of how lost they would feel if they had to move from their neighborhood, and the Exp(B) of 0.87 for Q29 indicates a 13% decrease in the odds for each unit increase in the measure of enjoying living in one's neighborhood (Q26 and Q29 were both measured with a 0-to-10 response scale.)

Moving to Block 3 we begin to see how a hierarchical model may impact the big picture. With three out of four variables in Block 3 reflecting a significant r, one might assume this block to have an even greater impact on the overall model than did Block 2. However, as we see in Table 1, none of the variables in Block 3 have a significant Exp(B), nor is

Block 3's overall Chi-square significant. The model Chi-square remains significant, and does increase to 38.63, but perhaps not as much of an increase as we may have predicted. We might hypothesize 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 2 variables. If the model were to be run by switching the order of Blocks 2 and 3, perhaps we would find the variables included in "Neighborhood Communication Involvement" to have significant Exp(B)'s.

Table 1 also reveals that the Hosmer & Lemeshow goodness-of-fit test (how well the model fits the data) was only found to be slightly significant ($p<0.10$) in Block 1, and not significant in Blocks 2 or 3, indicating a fairly good fit for the model overall. The -2LL for the full model is 466.44, which, given its high dependence on n, is better interpreted by Cox & Snell R^2 and Nagelkerke R^2. The Cox & Snell R^2 value of 0.10 in Block 3 indicates the independent variables in the full model explain approximately 10% of the variance in the dependent variable. This is further confirmed by the Nagelkerke R^2 of 0.13 for the full model, estimating 13% of the variance of the dependent variable is explained by the independent variables included in the overall model.

Table 2: Classification Results

		Final Predictions (Blocks 1-3)		
		Q83:Attended city council meetings		Percentage Predicted
Observed		No = 311	Yes = 72	Correctly
Q83:Attended city council meetings	No = 241	209	32	86.7%
	Yes = 142	102	40	28.2%
Overall Percentage				65.0%

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=383

n = 209 + 40 = 249

K = 2

Press' Q = $[383-(249*2)]^2 / 383(2-1)$

= $[383-498]^2 / 383$

= 13,225 / 383

Press' Q = 34.5 df =1 Critical chi-square at 0.001 level of significance = 10.83

As shown in Table 2, the model correctly classified 65% of the cases. The Press' Q calculation of 34.5 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.