

Statistics II

Aaron Kreider

May 2, 1999

Term Paper

Introduction

Generally possessing few financial resources, social movements rely upon volunteers for their sustenance. Therefore one of the large questions in social movement theory is what distinguishes the minority of people who become active, from the majority who remain uninvolved. By finding indicators that separate the two groups, one can better understand social movements, through understanding the make-up of their participants.

So this paper will look at the demographic, attitudinal, and other characteristics of activists compared to non-activists. The data comes from the General Social Survey (GSS -- 1996), as downloaded from the worldwide web (using: <http://csa.berkeley.edu:7502/archive.htm>). The 1996 GSS is particularly useful since it was the first time in over thirteen years that it included protest-related questions. Unfortunately there were not any questions regarding involvement of friends in protest-related activities, and this model could be better if it included measurements of how involvement in different networks affected one's chances of having protested.

The Dependent Variable

The question that I used for my dependent variable was included in a political module that was administered to 1332 people:

"692. And in the past five years how many times have you done each of the following to protest against a government action you strongly oppose? PLEASE CHECK ONE BOX ON EACH LINE...

B. Gone on a protest march or demonstration"

Possible answers included never, once, more than one, don't know, and no answer. Since relatively few people had either protested once (87) or more than once (only 31), I combined them and created a dichotomous dependent variable: PROT2. This also allowed me to use logistic regression. Some of the usefulness of the upcoming models is limited by the choice of this dependent variable. I lost information by combining the two categories of protestors. Also it is an imperfect operationalization of the concept of 'activist'. It is possible for activists to not attend protests or

demonstrations (Ex. an activist could be organizing conferences, doing speaking tours, and petitioning - without ever attending a protest), and more critically (and likely) it is possible for people to attend a protest just because there was a crowd and it looked interesting, instead of out of any activist motivation on their part. The GSS question says nothing about the specific type of protest, though since it is generally classified as 'against a government action', one can safely assume that it was a political issue (thus sports-related riots and other apolitical issues are excluded).

I tried constructing a protestor index out of the previous question and another one which asked whether in the past five years they had "Attended a public meeting organized to protest against the government". However there were substantial differences between the two groups. Also there were not any other measures that could be used to construct the index, so I felt it would be better to use the clear-cut question regarding protesting in the past five years.

Missing Data

During my first trial runs, about 40% of my cases were removed by list-wise deletion (since logistic regression does not do pair deletion) due to missing income data. By contrast other variables had only a small percent of missing data. I could not remove income since it, or more precisely its log, was an important predictor. So I regressed $\ln(\text{INCOME})$ on AGE, $\ln(\text{AGE})$, BLACK, CATHOLIC, CLASS, EDUCATION, JEWISH, MALE, MARRIED, PROTESTANT, WNCENT, and WSCEN (the last two are regions). See Figure 1. I chose the independent variables using backward OLS. I could have done a better job predicting income - but my intent was just to use a method better than mean substitution. Next, I substituted the predicted values for $\ln(\text{INCOME})$ when the data was missing. The subsequent results for logistic regressions were very similar whether I used list-wise deletion or substitution. I shall show those which used substitution since it gives a larger sample size, and best avoids the danger of the missing data having significant differences. Such differences may exist as only 6% of the cases with missing income data had protested (23 of 383), whereas 10.4% (95 of 912) of the cases without missing income data had protested. This difference was significant at a level between 0.05 and 0.06 (depending on the measurement method), which either occurred by chance or is evidence of a problem in the 1996 GSS.

Figure 1.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.566 ^a	.321	.317	.8236

a. Predictors: (Constant), Catholic, MARRIED, HIGHEST YEAR OF SCHOOL COMPLETED, WSCEN, MALE, West-North Central Region, JEWISH, AGE OF RESPONDENT, BLACK, CLASS, protestant, ln of age

b. Dependent Variable: ln of income

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	614.863	12	51.239	75.542	.000 ^a
	Residual	1301.619	1919	.678		
	Total	1916.482	1931			

a. Predictors: (Constant), Catholic, MARRIED, HIGHEST YEAR OF SCHOOL COMPLETED, WSCEN, MALE, West-North Central Region, JEWISH, AGE OF RESPONDENT, BLACK, CLASS, protestant, ln of age

b. Dependent Variable: ln of income

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.320	.876		-7.212	.000
	HIGHEST YEAR OF SCHOOL COMPLETED	8.108E-02	.008	.215	10.321	.000
	JEWISH	.265	.142	.037	1.873	.061
	MALE	.515	.038	.258	13.526	.000
	MARRIED	8.358E-02	.039	.042	2.116	.034
	West-North Central Region	-.174	.072	-.046	-2.401	.016
	WSCEN	-.171	.066	-.049	-2.607	.009
	protestant	.138	.053	.069	2.594	.010
	CLASS	.221	.033	.138	6.618	.000
	BLACK	8.336E-02	.058	.029	1.446	.148
	AGE OF RESPONDENT	-.116	.009	-1.455	-13.568	.000
	ln of age	5.152	.335	1.652	15.360	.000
	Catholic	.160	.059	.069	2.719	.007

a. Dependent Variable: ln of income

Basic overlay - Demographic Variables

The first model (Figure 2) consists of basic demographic variables such as race, sex, religion, age, marital status, and sexual orientation. For race, sex, and sexual orientation similar theories exist. One would either predict, using a Marxist approach, that the oppressed members of each group (thus non-whites, women, gays, lesbians, and bisexuals) would be more likely to protest than the members of the privileged groups. Also one could predict that since women, people of color, and queers have less power (a.k.a. 'The System' was built against them), one could predict that they would be more likely to use unconventional means to achieve their goals - and thus protest. A contradictory theory would have men, whites, and heterosexuals more likely to protest since they have more resources to mobilize. In addition one might predict that women (because of the gender stereotype of being 'passive' in contrast to the male 'active') and queers (because of their fear of being outed) might be less likely to protest due to social pressures. According to the initial results, none of these three factors are of much significance, though QUEER and WHITE will become more so as the model progresses. The odds of queers protesting is from 1.6 to 3 greater than heterosexuals (in the final model), and the odds of whites having protested are 30-40% less than non-whites.

Since the GSS does not ask direct questions about sexual orientation, I constructed the QUEER variable by combining a question on a person's sex with one on their sexual activity of the past five years. Thus QUEER includes lesbians (women who had sex exclusively with women), gays (men who had sex exclusively with men), and bisexuals (men or women who had sex with both sexes). I also constructed a CLOSET variable, made up of queers who were married. I thought that they would be less likely to protest since they would want to keep their identity hidden. However CLOSET was insignificant, likely since people could protest many issues not related to their sexuality without risk.

Religion is normally not a large factor in predicting activism, except for Jews who tend to be activist in significantly greater numbers. This result is often shown in other studies, and must be some characteristic of the Jewish religion and culture. This model matches those previous findings with Jews having greater odds by a factor of four (at this stage) to nine (in the final model).

It makes sense that married people would have less time due to greater responsibility and social attachments (to spouse, children, and home). Also they might be less likely to do deviant act, such as protesting, for the same reasons. The results (from Figure 3 on) show that MARRIED people had odds of having protested of six to seven times smaller.

Figure 2.

Number of cases included in the analysis: 1118

-2 Log Likelihood 687.220
 Goodness of Fit 1111.273
 Cox & Snell - R² .028
 Nagelkerke - R² .059

----- Variables in the Equation -----							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	1.5392	.5200	8.7615	1	.0031	.0970	4.6611
MALE	.2178	.2092	1.0842	1	.2978	.0000	1.2433
MARRIED	-.7718	.2195	12.3600	1	.0004	-.1200	.4622
QUEER	.5144	.4108	1.5681	1	.2105	.0000	1.6727
WHITE	-.4032	.2415	2.7866	1	.0951	-.0331	.6682
AGE	-.0070	.0076	.8677	1	.3516	.0000	.9930
PROTSANT	-.0661	.2756	.0574	1	.8106	.0000	.9361
CATH	-.0210	.3112	.0045	1	.9463	.0000	.9792
Constant	-1.4145	.3873	13.3364	1	.0003		

Adding Income, Education, and an Interaction Term

Next I build upon the base model by including INCOME and EDUCATION (Figure 3). One would expect that poor people are more likely to protest, whereas the rich are more likely to lobby and work within institutions to achieve their goals. Also according to a Marxist view, the poor would have more grievances as the government acts on behalf of the capitalist elite to help them exploit the workers. Education tends to have a liberalizing effect upon individuals (for instance as seen by the societal role of dissenting intellectuals), and when controlled for income, one would expect to find a positive correlation between it and having protested. To measure income, I first replaced the categorical values (of 1 to 22) with the mean value of that range. Next I took its log, since income varied from \$500 to \$100,000 and a thousand dollar income increase for a rich person should have less impact than for a poor one. The coefficients for ln (INCOME) and EDUCATION matched the predicted direction, though INCOME is not yet significant. According to the final model, someone earning \$74,000 would be half as likely to have protested as someone earning \$10,000. Also someone with an additional four years of education (Ex. college degree versus high school diploma) would have 50% greater odds.

Finally I included the effect of being a married male. It almost completely negates the effect of being married, probably as a married male is less subject to additional obligations such as children and housework. For instance in traditional male bread-winner marriages, men are freed from household chores in exchange for a transfer of income to their spouse and in some cases will have *more* time to engage in political protest.

Figure 3.

Number of cases included in the analysis: 1116

Interactions: INT_1 MALE by MARRIED

-2 Log Likelihood 649.842

Goodness of Fit 1117.977

Cox & Snell - R² .060

Nagelkerke - R² .126

----- Variables in the Equation -----							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	1.3414	.5442	6.0760	1	.0137	.0753	3.8243
MALE	-.3595	.2765	1.6911	1	.1935	.0000	.6980
MARRIED	-1.7460	.3717	22.0619	1	.0000	-.1671	.1745
QUEER	.6937	.4204	2.7229	1	.0989	.0317	2.0011
WHITE	-.4313	.2456	3.0843	1	.0790	-.0388	.6496
PROTSANT	.0607	.2817	.0465	1	.8293	.0000	1.0626
CATH	.1269	.3198	.1576	1	.6914	.0000	1.1353
EDUC	.2122	.0432	24.1214	1	.0000	.1755	1.2364
INCOMLN2	-.2174	.1229	3.1279	1	.0770	-.0396	.8046
INT_1	1.6792	.4740	12.5505	1	.0004	.1212	5.3613
Constant	-2.3574	1.1349	4.3144	1	.0378		

City Size and Region

One would expect that larger cities would be more likely to have protests as they are more likely to have the critical mass of activists necessary to hold them. And thus one would postulate that their inhabitants would be more likely to have protested than in small cities where protests might never happen. In addition, one might expect that certain regions would have more protests due to differences in regional social values and history. Neither size nor regions were found to be significant at this stage. In the final model people from the West-North-Central part of the U.S. have odds four times lower than people from the rest of the country.

Figure 4.

Number of cases included in the analysis: 1116

-2 Log Likelihood 634.884

Goodness of Fit 1305.368

Cox & Snell - R² .072

Nagelkerke - R² .152

----- Variables in the Equation -----							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	1.3308	.5540	5.7712	1	.0163	.0724	3.7843
MALE	-.3618	.2819	1.6478	1	.1993	.0000	.6964
MARRIED	-1.8064	.3784	22.7852	1	.0000	-.1701	.1643
QUEER	.7710	.4262	3.2723	1	.0705	.0421	2.1619
WHITE	-.3180	.2612	1.4822	1	.2234	.0000	.7276
PROTSANT	.2741	.2942	.8682	1	.3515	.0000	1.3154
CATH	.1028	.3247	.1003	1	.7515	.0000	1.1083
EDUC	.2188	.0442	24.5153	1	.0000	.1770	1.2446
INCOMLN2	-.2398	.1241	3.7330	1	.0533	-.0491	.7868
INT_1	1.7834	.4823	13.6750	1	.0002	.1275	5.9503
SIZE	.0001	7.772E-05	1.6898	1	.1936	.0000	1.0001
ENCENT	.0193	.4567	.0018	1	.9662	.0000	1.0195
ESCEN	-1.3902	.8204	2.8715	1	.0902	-.0348	.2490
MIDATL	.0165	.4822	.0012	1	.9726	.0000	1.0167
NEWENG	-.5649	.6592	.7343	1	.3915	.0000	.5684
PACIFIC	.3352	.4375	.5870	1	.4436	.0000	1.3982
SATLANT	-.2223	.4646	.2291	1	.6322	.0000	.8006
WNCENT	-.7862	.6096	1.6629	1	.1972	.0000	.4556
WSCEN	-.2524	.5268	.2295	1	.6319	.0000	.7770
Constant	-2.3638	1.1794	4.0170	1	.0450		

Adding Political Views

While the GSS question does not specify what people protested, since it was targeted at the government it is most likely that the protestor was either very liberal or very conservative. For what would moderates have to protest? In addition, even if the moderates did have things to protest, they would be less numerous than for people whose political position will more often diverge from that of the government. Thus I transformed a question that classified the person's political views on as scale from very liberal to very conservative, to a measure of political extremism which measures one's political distance from being a moderate. This first attitudinal measure came out very significant and the odds of people who are very liberal or very conservative having protested are four times greater. I tried using different measurements for political views (including splitting it up into six dummy variables), and the extremism one worked the best.

Figure 5.

Number of cases included in the analysis: 1068

-2 Log Likelihood	609.589
Goodness of Fit	1034.715
Cox & Snell - R ²	.077
Nagelkerke - R ²	.161

----- Variables in the Equation -----							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	1.3646	.5562	6.0194	1	.0141	.0760	3.9142
MALE	-.3689	.2835	1.6924	1	.1933	.0000	.6915
MARRIED	-1.7298	.3758	21.1825	1	.0000	-.1661	.1773

QUEER	.7940	.4309	3.3947	1	.0654	.0448	2.2122
WHITE	-.4769	.2542	3.5198	1	.0606	-.0468	.6207
PROTSANT	.1362	.2938	.2149	1	.6429	.0000	1.1459
CATH	.2540	.3313	.5881	1	.4432	.0000	1.2892
EDUC	.2164	.0450	23.1396	1	.0000	.1744	1.2416
INCOMLN2	-.2387	.1277	3.4955	1	.0615	-.0464	.7876
INT_1	1.5920	.4823	10.8972	1	.0010	.1131	4.9137
POLEXTRM	.3986	.1152	11.9732	1	.0005	.1198	1.4898
WNCENT	-.6895	.4934	1.9523	1	.1623	.0000	.5018
Constant	-2.6408	1.1896	4.9280	1	.0264		

Adding Protest Meetings

The 1996 GSS also included a question as to whether the person had attended one, many, or no public meetings to protest against the government. It is logical that such persons would be more likely to participate in demonstrations. I split the variable into two dummies, whose significance show a large positive non-linear effect. People who attended multiple protest meetings had their odds of having demonstrated increased by over 31 times, and just by attending one meeting their odds increased by over 13 times. Adding these variables substantially improved the amount of variance the model explained, with both measures of R squared doubling.

Figure 6.

Number of cases included in the analysis: 1067

Variable(s) Entered on this Step

P15A1 Attended many protest meetings
P15A2 Attended one protest meeting

-2 Log Likelihood 461.927
Goodness of Fit 1240.084
Cox & Snell - R² .193
Nagelkerke - R² .405

----- Variables in the Equation -----							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	2.0705	.6774	9.3424	1	.0022	.1031	7.9284
MALE	-.3907	.3321	1.3846	1	.2393	.0000	.6766
MARRIED	-2.0484	.4310	22.5829	1	.0000	-.1726	.1289
QUEER	.7057	.5074	1.9344	1	.1643	.0000	2.0252
WHITE	-.5006	.2991	2.8010	1	.0942	-.0341	.6061
PROTSANT	.5647	.3571	2.5003	1	.1138	.0269	1.7590
CATH	.5429	.3908	1.9301	1	.1648	.0000	1.7211
EDUC	.1463	.0526	7.7371	1	.0054	.0911	1.1576
INCOMLN2	-.3438	.1481	5.3865	1	.0203	-.0700	.7091
INT_1	1.8184	.5581	10.6173	1	.0011	.1117	6.1619
POLEXTRM	.3207	.1303	6.0616	1	.0138	.0767	1.3781
WNCENT	-1.3014	.6171	4.4477	1	.0349	-.0595	.2722
P15A1	3.4552	.3673	88.5033	1	.0000	.3539	31.6631
P15A2	2.6096	.2890	81.5461	1	.0000	.3394	13.5936
Constant	-1.7101	1.3949	1.5030	1	.2202		

Adding Attitude towards protesting

The final model includes the response to the question:

"Would you or would you not do any of the following to protest against a government action you strongly opposed? PLEASE CHECK ONE BOX ON EACH LINE.
B. Go on a protest march or demonstration."

The responses of 'definitely would' and 'probably would' were grouped together, as were 'probably would not' and 'definitely would not'. It's clear that people who support protesting as a tactic would be more likely to do so themselves, however there is also the problem of reverse causality. It might easily be that the fact that someone who protested during the past five years changed his or her answer on the question. The crosstabulation shows that there are some people who had protested who advise against it, though the correlation is clear. And the people who had an attitude favoring protest had 4.5 times greater odds of having done it themselves.

Figure 7.

**Would you Go on a protest march-demonstration * Protested in past five years
Crosstabulation**

Count		Protested in past five years		Total
		0	1	
Would you Go on a protest	1	207	63	270
march-demonstration	2	263	40	303
	3	352	12	364
	4	298	2	300
Total		1120	117	1237

Figure 8.

Number of cases included in the analysis: 1029
P3 would or definitely would protest

-2 Log Likelihood 436.017
Goodness of Fit 1067.826
Cox & Snell - R² .213
Nagelkerke - R² .439

Classification Table for PROT2

The Cut Value is .50

		Predicted			Percent Correct
		0	I	1	
Observed	0	I 904	I 19	I	97.94%
	1	I 61	I 45	I	42.45%
Overall				92.23%	

Variables in the Equation							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
JEWISH	2.2265	.6662	11.1688	1	.0008	.1159	9.2678
MALE	-.2413	.3423	.4970	1	.4808	.0000	.7856
MARRIED	-1.8017	.4352	17.1370	1	.0000	-.1489	.1650
QUEER	1.1838	.5438	4.7389	1	.0295	.0633	3.2669
WHITE	-.5436	.3057	3.1616	1	.0754	-.0413	.5807
PROTSANT	.6318	.3599	3.0809	1	.0792	.0398	1.8810
CATH	.5677	.3948	2.0675	1	.1505	.0099	1.7642
EDUC	.1066	.0544	3.8462	1	.0499	.0520	1.1125
INCOMLN2	-.3585	.1536	5.4476	1	.0196	-.0711	.6987
INT_1	1.7109	.5669	9.1068	1	.0025	.1020	5.5337
POLEXTRM	.2773	.1333	4.3287	1	.0375	.0584	1.3196
P15A1	3.0632	.3702	68.4802	1	.0000	.3121	21.3956
P15A2	2.3413	.2957	62.6772	1	.0000	.2982	10.3951
P3	1.5191	.3544	18.3684	1	.0000	.1549	4.5682
WNCENT	-1.3975	.6179	5.1157	1	.0237	-.0676	.2472
Constant	-2.0897	1.4642	2.0370	1	.1535		

Conclusion

The final model manages to predict a decent number of the protestors. Also the estimated coefficients make theoretical sense, confirming the theory that members of traditionally oppressed groups are more likely to demonstrate. Protestors are more likely to be Jewish, female, queer, well educated, unmarried, poor, a political 'extremist', to have previously attended protest meetings, and to have an attitude in favor of demonstrations. The major problem with my approach is not knowing the direction of causality. This is especially a problem for some of the most significant variables, like whether the person would attend a protest and whether they had attended a protest meeting. Without these variables, while the model is more likely to be correct in its causal direction (since one is less likely to alter one's religion or region, because one attended a couple protests), it explains a lot less variance. This problem could be solved by gathering data at two points of time, using the early data as background to predict whether people protested.