DATA ANALYSIS USING SETUPS AND SPSS: AMERICAN VOTING BEHAVIOR IN PRESIDENTIAL ELECTIONS 1972-2008

The original *SETUPS: AMERICAN VOTING BEHAVIOR IN PRESIDENTIAL ELECTIONS 1972-1992* module was based on combined (or "pooled cross-section") data from the 1972 through 1992 *American National Election Studies* (ANES). ANES studies have been held in conjunction with every Presidential election and most (off-year) Congressional elections since 1952. A large portion of political science knowledge concerning U.S. electoral behavior is derived from this series of studies. For a brief description of these studies, see the *SETUPS: ANES 1972-2008 DATA AND CODEBOOK* handout. As explained in that handout, the data available to POLI 300 students has now been extended through the 1996, 2000, 2004, and 2008 elections.

Each American National Election Study is a survey of approximately two thousand randomly selected respondents who collectively constitute (we can confidently expect, for reasons to be discussed in class) a representative sample of the American voting-age population at the time. Since *ten* national samples are combined here, the *total* number of respondents is approximately 20,000. In presidential election years, survey respondents are interviewed both before and after the November election. The *SETUPS* version of this data is considerably "condensed," in that it includes data (i) only for respondents who were successfully interviewed both before and after the election and (ii) only for a subset of the questions asked on each of the (very long) questionnaires. Moreover, possible responses to many questions have been simplified or combined and open-ended questions have been recoded in closed format.

Each category of information (vote for President, party identification, opinion on abortion, age, etc.) elicited from respondents (by means of a survey question or combination of questions) is an example of a *variable*. Each possible answer to a given question (or combination of questions) constituting a variable is called a *value* of the variable. (Thus values of the variable "How did you vote for President?" are "Bush," "Gore," "Nader," etc.; values of the variable "What is your party identification?" are "strong Democrat," "Independent," etc.).

In order to compactly record the very large amount of data that is collected in such surveys, data is *coded* in *numerical* form. This means that: (i) each respondent (or "case") is assigned an ID number; (ii) each variable is assigned an essentially numerical name; and, in particular, (iii) each value of each variable is assigned a numerical code. Thus the *SETUPS* data is recorded as an enormous rectangular *data array* of numbers (or *spreadsheet*). The four "corners" of the *SETUPS* data are shown Figure 1 below.

FIGURE 1. SETUPS DATA ARRAY (SPREADSHEET)

		Variables							
	CASE ID	V01	V02	V03	V04	.....	V69	V70	WT1
	1	1972	9	2	9	.....	3	3	1.000000
C	2	1972	9	1	2	.....	3	3	1.000000
a	3	1972	9	9	9	.....	3	3	1.000000
s	4	1972	9	1	2	.....	2	3	1.000000
e	.	.	.	.	.	.....	.	.	
s	.	.	.	.	.	.....	.	.	
	19972	2008	1	1	2	.....	2	9	1.016700
	19973	2008	2	2	9	.....	2	9	1.809800

As indicated in Figure 1, the SETUPS data array has 72 columns (one for each variable V01 through V70 plus CASEID and WT1) and 19,973 rows (one for each respondent). If we look across any *row* of this array of numbers, we see the (coded) value of each of the 70 variables V01 through V70 for a given (anonymous) respondent — in effect, how a given respondent answered each of 70 questions. If we look down any *column* of this array, we see how a given variable takes on different values from respondent to respondent (from case to case) — in effect, how a given question was answered by each of the respondents.

Of course, we can't do this in a meaningful way unless we can *decode* this numerical information. That is, in order to analyze and interpret the results of a survey, we must be provided with a *codebook*, in addition to the coded data. The Codebook for this data is provided in Handout #1B; Handout #1A provides an explanation of how to use the Codebook.

The Codebook tells us the substantive nature of each variable V01 through V70, and the substantive nature of each coded value for each variable. Using the Codebook in conjunction with the (partial) data array in Figure 1, we can see that respondent 1 did not vote (is coded 2 on V03), accordingly did not vote for a Presidential candidate (is coded 9 or "missing data" on V04), and so forth. Respondent 2, on the other hand, did vote and voted for the Republican candidate. Looking down the V03 column, we see that 1 failed to vote, 2 did vote, 3 is missing data, 4 did vote, and so forth.

Variable V01 indicates the year in which the respondent was interviewed and is the only variable other than Case ID and WT1 whose values have *not* been coded — instead the actual election year (or number or numerical weight) is recorded. Since the ten elections surveys are accumulated in chronological order, 1972 appears in the V01 column for the first 2706 cases (the exceptionally large size of the 1972 NES sample). V02 is REGISTERED TO VOTE? but, as the Codebook notes, this data was not available for 1972, so 9 (NA or "missing data") appears in the V02 column for the first 2706 cases.

Of course, given such a large data array (19,973 respondents times 70 variables equals 1,398,110 recorded values), it would be extraordinarily time-consuming and tedious to tabulate and analyze the survey data by hand.[1] It is far quicker and more convenient to use a machine — a counter-sorter machine many decades ago, a mainframe computer a couple of decades ago, a PC today — to do this processing for us. Thus you are being provided with access to a *computer data file* that contains the full data array; the file also contains *labels* (descriptive names) for all the

[1]The combined size of the ten NES samples is 19,973 respondents. Because of complexities pertaining to sampling procedures and contacting of respondents, in some years respondents must be *weighted* unequally in order to produce a representative sample. The final variable WT1 in the data array specifies the appropriate weighting. (As Figure 1 suggests, weighting is required for the 2008 but not 1972 data.) Because of weighting, it normally appears in tables that there are about 20,583 respondents (including missing data). Such weighting also means that, while case counts are always displayed as whole numbers, they are subject to *rounding error*, like percentages (usually displayed to the nearest tenth of a percentage point), so you will find that case counts sometimes appear not to add up properly. A further complication arises because the ten NES samples are not the same size. (In particular, the 1972 and 1976 samples are considerably larger than the later ones.) For some purposes, it might be appropriate to weight cases so that each of the ten election samples accounts for a 10% share of the total weighted sample. However, the SETUPS data has not been weighted in this fashion, since we almost always analyze data separately for each election year.

variables and their values (matching those shown in the Codebook). You are also being provided with access to a *computer program* called *SPSS (Statistical Package for the Social Sciences)* by which you can analyze this data.

Handout #1D on *USING SETUPS 1972-2008 ANES DATA AND SPSS FOR WINDOWS* provides you with the "nuts and bolts" information you need to open this data file and perform simple SPSS analyses in any UMBC PC lab. The remainder of this handout provides examples of the kinds of things you can do once you master these "nuts and bolts."

You will use SPSS to generate *tables* classifying the survey data and displaying case counts or percentage frequencies.

The simplest sort of table is a *frequency distribution* of a *single variable*. Such a table simply shows how many respondents (*absolute frequencies*), or what percent of respondents (*relative frequencies*), have each value on a given variable. Let us consider a couple of particular examples.

Recorded turnout in Presidential elections from 1972 through 2008 has ranged from about 49% (in 1996) to 62% (in 2008). We can see what the corresponding percentages are in our sample of respondents by having SPSS construct a frequency distribution for variable V03 (VOTED IN ELECTION). The result is shown in Table 1 (which is actual SPSS output but sightly edited [in particular, the numerical value codes have been added] — the format can be modified in various ways).

TABLE 1: FREQUENCY DISTRIBUTION OF V03 (VOTED IN ELECTION)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1 voted	13138	63.8	73.7	73.7
	2 did not vote	4695	22.8	26.3	100.0
	Total	17833	86.6	100.0	
Missing	9 NA	2751	13.4		
	Total	20583	100.0		

Bear in mind that the computer did nothing magical — it simply (1) *read down the V03 column in the data array*, (2) *tallied up the number of 1's, 2's, and 9's in the column, (3) calculated the corresponding percentages*, and (4) *printed the results* (together with appropriate labels).

Table 1 shows both the *variable number* (V03) and the *variable label* (VOTED IN ELECTION) and both the *value codes* (1, 2, and 9) and the *value labels* ("voted," "did not vote," and NA [*missing data*]) and, for each value, shows: (i) *absolute frequencies* or *case counts* (in the "Frequency" column), i.e., the actual number of cases having each value; (ii) *relative frequencies* (in the "Percent" column), i.e., the absolute frequencies as *percentages* of *all* 20,583 cases; and (iii) *adjusted relative frequencies* (in the "Valid Percent" column), i.e., the absolute frequency as a percent of all 17,833 cases *after excluding missing data*, i.e., excluding all cases coded as "9"or NA

("not applicable/not ascertained").[2] (It also shows (iv) *cumulative frequencies*, which are unhelpful or make no sense in this context, so we will not discuss them further here.) Ordinarily we are unlikely to be interested in the ("unadjusted" relative frequency) entries in the "Percent" column, because these relative frequencies are calculated over *all* respondents in the survey, including the missing data cases that we know nothing about. We are more likely to be interested in the entries in the "Valid Percent" column, based *only* on respondents who gave useful answered to the relevant question. Indeed, most tables in articles and books do not display missing data at all.

What we see looking at the "Valid Percent" column is that reported turnout in our pooled sample is much higher than what we have actually seen in recent Presidential elections. Partly this is because some people do not answer this question truthfully, but other more subtle factors contribute importantly to this *upward bias* in survey results (and will be discussed in class later).

To take another example, the commonly reported division of the popular vote in the 1992 Presidential election was about 43% for Bill Clinton, 38% for George Bush, and 19% for Ross Perot. Again can see what the corresponding percentages are in our sample of respondents by having SPSS construct a frequency distribution for variable V04 (PRESIDENTIAL VOTE) *for 1992 respondents only*.

TABLE 2 — FREQUENCY DISTRIBUTION OF V04 (PRESIDENTIAL VOTE) FOR 1992 ONLY

		Frequency	Percent	Valid Percent	Cum Percent
Valid	1 Dem	793	31.9	47.7	47.7
	2 Rep	562	22.6	33.9	81.6
	3 Other	306	12.3	18.4	100.0
	Total	1661	66.8	100.0	
Missing	9 NA	827	33.2		
Total		2488	100.0		

In this case, the computer did *two* things. First, it sorted through all the cases and "filtered out" all cases *except respondents in the 1992 survey* (i.e., all except the 2488 cases with a "1992" in the V01 column of the data array). Then, with the remaining cases (the 2488 respondents only) after the filtering operation, it read down the V04 column in the data array and tallied up the number of 1's, 2's, 3's, and 9's in the column, and calculated the percentages.[3]

The entries in the "Percent" column of Table 2 deviate greatly from the actual election results. But this is because these relative frequencies are calculated over *all* respondents in the survey, including the *missing data* (particularly including respondents who previously reported

---

[2] Settings in the data file tell SPSS that code 9 represents missing data.

[3] Note that Presidential candidates are labeled not by name but by party, since the same labels must apply across the entire 1972-2004 period.

[V03] that they did not vote at all). However, the entries in the "Valid Percent" column, based only on respondents who reported voting, quite closely match the known election results (though support for the winner is somewhat exaggerated — a common phenomenon in surveys).

Note that Table 1 pools together all respondents in all surveys from 1972 through 2008. Given a pooled cross-section like this data, it often is not very enlightening to look at all cases pooled together like this (especially given that the ten election year samples are not the same size). We are more likely to want to examine one cross-section (respondents in one election year) only, in the manner of Table 2. But what may be even more enlightening is to conduct *longitudinal* (over time) *analysis* and look at *all* the cross-sections (election years) *in turn* and make comparisons among them. This could be accomplished by having the computer do what it did for 1992 in Table 2 (with respect to Presidential vote) for each election year in turn. But since V01 (YEAR OF SURVEY) is just another variable, we can *crosstabulate* (this procedure is discussed in more detail below) the variable of interest with V01 and produce a table like the following. (This table has been reformatted in a compact fashion to look as it might appear in an article or book, showing only adjusted relative frequencies plus the number of [non-missing] cases for each year. We could make this table even more compact by deleting the "Didn't vote" and "100%" rows, since (with missing data excluded) always "Didn't vote" = 100% – "Voted".)

TABLE 3. ANES VOTING TURNOUT FROM 1972 THROUGH 2004

Voted	1972	1976	1980	1984	1988	1992	1996	2000	2004	2008
Yes	72.8	71.6	71.4	73.6	69.7	75.1	77.0	72.1	80.0	77.6
No	27.2	28.4	28.6	26.4	30.3	24.9	23.0	27.9	20.0	22.4
	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	(n=2283)	(n=2403)	(n=1407)	(n=1989)	(n=1773)	(n=2256)	(n=1521)	(n=1551)	(n=535)	(n=2113)

In the remaining examples, we will focus on the 1992 cross-section only (filtering out all other respondents in the manner of Table 2). One issue that clearly divided the two major candidates and parties in 1992 (especially) was abortion. We can ask SPSS to produce a frequency table for V45 (ABORTION).

TABLE 4: FREQUENCY DISTRIBUTION OF 1992 ABORTION OPINION (V45)

		Frequency	Percent	Valid Percent	Cum Percent
Valid	**1 Never permit**	257	10.3	10.7	10.7
	2 For rape, etc.	681	27.4	28.3	39.0
	3 Need established	344	13.8	14.3	53.2
	4 Always permit	1126	45.3	46.8	100.0
	Total	2408	96.8	100.0	
Missing	**9 NA**	80	3.2		
Total		2488	100.0		

We now have frequency distributions of both 1992 PRESIDENTIAL VOTE (Table 2) and 1992 ABORTION OPINION (Table 4). We would probably expect that most people with more "pro-choice" views on abortion voted for Clinton (or perhaps Perot) rather than Bush, while most of those with more "pro-life" views voted for Bush. The preponderance of "pro-choice"views on abortion in the electorate may thus help account for Clinton's victory. But just looking at these two frequency distributions in Tables 2 and 4 provides *no evidence* for or against the hypothesis that such an association between abortion opinion and voting behavior exists.

What we must do instead is create a somewhat more complicated kind of *two-variable* table called a *crosstabulation*. Such a table shows, for all cases that have a given value on one variable, their frequency distribution with respect to the other variable. Let us have SPSS create a cross-tabulation of V04 and V45 (for 1992 only) to test the expectations developed above. Here is the result.

TABLE 5A: CROSSTABULATION OF PRESIDENTIAL VOTE (V04) BY ABORTION OPINION (V45)
(Case Counts [Absolute Frequencies])

		ABORTION					Total
		1 Never	2 Rarely	3 Need	4 Always	9 NA	
PRES	1 Dem	56	165	93	462	16	792
	2 Rep	66	202	100	176	19	563
	3 Other	19	69	45	164	9	306
	9 NA	116	246	106	323	36	827
Total		257	682	344	1125	80	2488

This table shows absolute frequencies only, not percentages. The missing data row and column are shaded. Notice that the row and column totals are simply the absolute frequencies for V04 (Table 2) and V45 (Table 4) respectively.[4] (Since they appear at the right and bottom "margins" of the crosstabulation, they are sometimes called *marginal frequencies* or simply *marginals*.) This is the information we can get from the separate frequency distributions; what we can't get from frequency distributions themselves is information about how the cases are distributed over the *interior cells* of the table. For this we need to crosstabulate the raw data, as has been done in Table 5A.

Again, we should consider what the computer did in constructing this crosstabulation. It looked down the V04 and V45 columns of the 1992 portion of the data array *simultaneously* and tallied up the different *combinations of values* it found. For example, it found that 56 respondents had the 1-1 (Clinton-Never Permitted) combination, 66 had the 2-1 (Bush-Never Permitted) combination, and so forth.

[4] The small discrepancies result from the rounding of weighted case counts, as discussed in footnote 1.

It appears that our general expectations are borne out, but the pattern can be made more apparent by: (i) excluding missing data, and (ii) calculating adjusted relative frequencies (percentages). But, since we have *two* variables, there are several ways to calculate percentages. This is illustrated by the following "panels" of the same crosstabulation. SPSS can calculate and display any or all such percentages, along with the absolute frequencies.

TABLE 5B: CROSSTABULATION OF PRESIDENTIAL VOTE BY ABORTION OPINION

(Row Percentages)

			ABORTION				Total
			1 Never	2 Rarely	3 Need	4 Always	
PRES VOTE	1 Dem	Count	56	165	93	462	776
		% within VOTE	7.2%	21.3%	12.0%	59.5%	100.0%
	2 Rep	Count	66	202	100	176	544
		% within VOTE	12.1%	37.1%	18.4%	32.4%	100.0%
	3 Other	Count	19	69	45	164	297
		% within VOTE	6.4%	23.2%	15.2%	55.2%	100.0%
Total		Count	141	436	238	802	1617
		% within VOTE	8.7%	27.0%	14.7%	49.6%	100.0%

The percentages in Table 5C have been calculated by taking each cell entry in Table 5 as a percentage of its *row total* (after excluding missing data, i.e., they are adjusted relative frequencies). These percentages tell us, of all respondents who have a given (non-missing) value on the row variable, what percent have a particular value with respect to the column variable. For example, in this in case we are told that, of all 544 respondents who voted for Bush, 32.4% (= 176/544) believe abortion should always be permitted. More generally, we see that Clinton and Perot voters had quite similar distributions of opinions on abortion, since the *row* ("% within VOTE") *percentages* are very similar in the "1 Dem" and "3 Other" rows, and that both groups of voters leaned distinctly in the pro-choice direction. In contrast, while the Bush voters (in the "2 Rep" row) are also preponderantly "pro-choice," they are relatively more "pro-life" than the other voters.

TABLE 5C: CROSSTABULATION OF PRESIDENTIAL VOTE BY ABORTION OPINION

(Column Percentages)

			ABORTION				Total
			1 Never	2 Rarely	3 Need	4 Always	
PRES VOTE	1 Dem	Count	56	165	93	462	776
		% within ABORTION	39.7%	37.8%	39.1%	57.6%	48.0%
	2 Rep	Count	66	202	100	176	544
		% within ABORTION	46.8%	46.3%	42.0%	21.9%	33.6%
	3 Other	Count	19	69	45	164	297
		% within ABORTION	13.5%	15.8%	18.9%	20.4%	18.4%
Total		Count	141	436	238	802	1617
		% within ABORTION	100.0%	100.0%	100.0%	100.0%	100.0%

The percentages in Table 5C are calculated by taking each cell entry in Table 5A as a percentage of its *column total* (after excluding missing data). Thus such percentages tell us, of all respondents who have a given (non-missing) value with respect to the column variable, what percent have a particular value on the row variable. For example, in this case we are told that, of all 141 respondents who believe abortion should never be permitted, 46.8% (= 66/141) voted for Bush. forth. More generally, we see that voters in the first three (more restrictive) abortion opinion categories all have quite similar distributions of Presidential voting, since the *column* ("% within ABORTION") *percentages* are quite similar in the "1 Never," "2 Rarely," and "3 Other" rows, and that such voters preponderantly supported Bush with Clinton close behind. In contrast, the most "pro-choice" voters (in the "4 Always" column) strongly supported Clinton and gave Bush hardly more support than Perot.

TABLE 5D: CROSSTABULATION OF PRESIDENTIAL VOTE BY ABORTION OPINION

(Total Percentages)

			ABORTION				Total
			1 Never	2 Rarely	3 Need	4 Always	
PRES VOTE	1 Dem	Count	56	165	93	462	776
		% of Total	3.5%	10.2%	5.8%	28.6%	48.0%
	2 Rep	Count	66	202	100	176	544
		% of Total	4.1%	12.5%	6.2%	10.9%	33.6%
	3 Other	Count	19	69	45	164	297
		% of Total	1.2%	4.3%	2.8%	10.1%	18.4%
Total		Count	141	436	238	802	1617
		% of Total	8.7%	27.0%	14.7%	49.6%	100.0%

The percentages in Table 5C are calculated by taking each cell entry in Table 5A as a percentage of the *grand total* in the table (after excluding missing data). Thus such percentages tell us, of *all* 1617 respondents in the entire table (in all rows and all columns), what percent have a particular *combination of values* with respect to the two variables. For example, in this case we are told that, of all respondents (who voted in the Presidential election and have an opinion on abortion), 28.6% (= 462/1617) believe abortion should always be permitted and also voted for Clinton.

In fact, SPSS can produce all four panels (Tables 5A, 5B, 5C, and 5D) in a single table like the following.

TABLE 5: CROSSTABULATION OF PRESIDENTIAL VOTE BY ABORTION OPINION (All Percentages)

			ABORTION				Total
			1 Never	2 Rarely	3 Need	4 Always	
PRES VOTE	1 Dem	Count	56	165	93	462	776
		% within VOTE	7.2%	21.3%	12.0%	59.5%	100.0%
		% within ABORTION	39.7%	37.8%	39.1%	57.6%	48.0%
		% of Total	3.5%	10.2%	5.8%	28.6%	48.0%
	2 Rep	Count	66	202	100	176	544
		% within VOTE	12.1%	37.1%	18.4%	32.4%	100.0%
		% within ABORTION	46.8%	46.3%	42.0%	21.9%	33.6%
		% of Total	4.1%	12.5%	6.2%	10.9%	33.6%
	3 Other	Count	19	69	45	164	297
		% within VOTE	6.4%	23.2%	15.2%	55.2%	100.0%
		% within ABORTION	13.5%	15.8%	18.9%	20.4%	18.4%
		% of Total	1.2%	4.3%	2.8%	10.1%	18.4%
Total		Count	141	436	238	802	1617
		% within VOTE	8.7%	27.0%	14.7%	49.6%	100.0%
		% within ABORTION	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	8.7%	27.0%	14.7%	49.6%	100.0%

Notice that Table 5, like Tables 5B, 5C, and 5D, excludes the "missing value" row and column shown (shaded row and column) in Table 5A. As a result, the total number of cases shown in these tables is:

	2488	(original number of cases in Table 5A)
minus	827	(missing on V04)
minus	80	(missing on V45)
	1581	
plus	36	(missing in both V04 and V45 and double-counted in the subtraction above)
	1617	

We now consider the possible impact of a *third variable* on the relationship between vote and abortion opinion. Let us consider the third variable AGE OF RESPONDENT (V60). Before the early 1970s, abortion was generally illegal and uncommon (or at least hidden from view and not talked about much). Therefore, we might expect that older voters, who came of age in less permissive times, would have more restrictive views concerning abortion than younger voters. To test this expectation, we can ask SPSS to crosstabulate V45 with V60.

TABLE 6: CROSSTABULATION OF ABORTION OPINION (V45) BY AGE (V60)

(Column Percentages)

		AGE						Total
		1 17-24	2 25-34	3 35-44	4 45-54	5 55-64	6 65-99	
ABOR-TION	1 Never	10.3%	7.1%	11.1%	9.6%	12.8%	14.4%	10.6%
	2 Rarely	32.3%	27.6%	21.5%	28.8%	36.4%	29.8%	28.3%
	3 Need	10.8%	13.5%	12.1%	18.1%	16.4%	15.0%	14.3%
	4 Always	46.6%	51.8%	55.3%	43.5%	34.4%	40.8%	46.8%
Total		223	591	503	375	250	466	2408
		100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

We see that the hypothesis receives only modest support. A more sophisticated hypothesis requires us to examine the *three* variables V04, V45, and V60 *simultaneously*. We might expect that the abortion issue would be highly salient to younger voters, both because it can affect them in a direct and personal way and because younger voters have come of age and acquired their political attitudes in an era during which the abortion issue has been prominently debated and — perhaps more clearly than any other single issue — has divided the political parties. On the other hand, the abortion issue does not so directly and personally affect older voters; perhaps more importantly, older voters came of age and acquired their political attitudes in earlier eras when abortion was not at all an issue in elections and when the political parties were more clearly divided on other quite different issues (basically pro/anti-New Deal and, a more recently, pro/anti-civil rights).

What we can do is to crosstabulate V04 and V45 while *controlling* for age (V60). To do this, we can have SPSS *recode* all respondents into three broad age categories: "younger" or 17-34 (V60 code categories 1 and 2), "middle aged" or 35-54 (code categories 3 and 4), and "older" or 55+ (code categories 5 and 6) and construct a *separate crosstabulation for each age category*. (Note that, within a single cross-section like this, controlling for age is equivalent to controlling for *generation* — that is, for when the respondents were born. In generational terms, the "younger" portion of the 1992 electorate was composed of voters born between 1958 and 1974, the "middle aged" category was composed of voters born between 1938 and 1957, and the "older" category was

composed of voters born in 1937 or earlier. But if we pooled the cross-sections together, the same age categories would be associated with different birth dates in different cross-sections.)

TABLE 7: CROSSTABULATION OF PRESIDENTIAL VOTE BY ABORTION OPINION CONTROLLING FOR AGE CATEGORY (Column Percentages)

			ABORTION				Total
AGE CATEGORY			1 Never	2 Rarely	3 Need	4 Always	
YOUNGER	PRES VOTE	1 Dem	32.3%	32.3%	36.2%	53.8%	44.6%
		2 Rep	51.6%	45.7%	41.4%	20.1%	31.5%
		3 Other	16.1%	22.0%	22.4%	26.1%	24.0%
	Total		31	127	58	264	480
			100.0%	100.0%	100.0%	100.0%	100.0%
MIDDLE AGED	PRES VOTE	1 Dem	28.3%	31.5%	40.6%	60.1%	47.8%
		2 Rep	56.6%	48.6%	40.6%	21.1%	33.5%
		3 Other	15.1%	19.9%	18.8%	18.7%	18.7%
	Total		53	146	96	331	626
			100.0%	100.0%	100.0%	100.0%	100.0%
OLDER	PRES VOTE	1 Dem	54.4%	47.6%	40.5%	58.7%	51.7%
		2 Rep	35.1%	44.5%	42.9%	25.5%	35.5%
		3 Other	10.5%	7.9%	16.7%	15.9%	12.9%
	Total		57	164	84	208	513
			100.0%	100.0%	100.0%	100.0%	100.0%

Based on this analysis, our hypothesis receives considerable support. In 1992, abortion opinion was substantially related to the way younger citizens voted, but was hardly related at all to the way older citizens voted. If there is a surprise, it is that the abortion issue if anything appears to have more influence on middle-aged voters than on younger ones.

We could further extend this kind of analysis by making use of the pooled cross-section and repeating it for other election years to see whether the pattern changes from 1972 to 2008.

Hopefully, these examples have suggested how you can develop hypotheses about American voting behavior and then test your hypotheses empirically by using SPSS to analyze the SETUPS

1972-2000 survey data. As previously noted, the accompanying handout on *USING SETUPS 1972-2000 NES DATA AND SPSS FOR WINDOWS* provides you with the "nuts and bolts" information you need to open this data file and perform simple SPSS analyses in any UMBC PC lab. Several of the POLI 300 Problem Sets will ask you to exactly this. In the event you feel sufficiently ambitious and empowered, this data and the SPSS software will remain available for your use beyond POLI 300, e.g., for research projects in other courses, for individual study projects, or for a departmental honors research project.