# SETUPS: ANES 1972-2008 DATA

## 1. *The American National Election Studies (ANES)*

In the Fall of 1948, the University of Michigan Survey Research Center (SRC) conducted a large-scale and carefully designed national survey. The interviews took place over a period of a month or two that happened to coincide with the run-up to the November 1948 Presidential election. Though the survey questionnaire focused on consumer confidence and expectations, it also contained a small number of questions pertaining political preferences and voting intentions with respect to the election. When the election turned out to be a stunning upset — with President Harry S. Truman defeating the Republican candidate Governor Thomas E. Dewey, the predicted winner in almost all commercial (Gallup, etc.) polls — the Michigan researchers took a closer look at the political data in their survey, which apparently could help account for the election outcome. Among other things, the SRC researchers compared the voting intentions expressed in their earlier and later interviews and discovered a substantial surge in support for Truman in the interviews closest to election day — a surge which the commercial pollsters, having already projected a Dewey victory, largely missed.

As a result of this fortuitous experience, the SRC secured resources to plan and carry out a national survey in Fall 1952 that would focus specifically on political, not economic, questions. Over the following years, the Michigan researchers carried out similar surveys in conjunction with Presidential and (off-year) Congressional elections. The University of Michigan created a new research unit separate from the SRC called the Center for Political Studies (CPS) to carry out these election studies, which came to be known informally as the "Michigan studies."

In the 1970s, the CPS secured (more or less) permanent funding from the National Science Foundation (a federal agency that funds much scientific research) and the former "Michigan studies" became known officially as the *National Election Studies* (NES) — or more specifically (as political scientists in other countries developed their own national election studies) as the *American National Election Studies* (ANES). Their design is shaped by a Board of Overseers and research staff representing the political science discipline as a whole, not simply researchers at Michigan, and ANES survey data is made available to all researchers simultaneously (about six months after an election).

The American National Election Studies now cover 60 years of American national elections. In addition to the data sets produced by each of the individual election studies, the ANES has created a *Cumulative Data File* which pools data from all the individual election studies into common data set that facilitates "longitudinal" (over time) analysis and allows researchers to examine electoral factors that may change from election to election and trends that may be evident only over long periods of time. (Recall the "national" charts shown in class displaying trends in party identification, ideology, abortion opinions, etc., over time. The corresponding "student" charts were based on a similar cumulative data file of student survey data.)

## 2. *The Interuniversity Consortium for Political and Social Research (ICPSR)*

As the Michigan SRC researchers completed their early election studies, they needed to find a way to make their valuable data available to other researchers (as required by the funding agencies that provided financial support for the surveys, and as they wanted to do in any event). In 1962, the *Interuniversity Consortium for Political and Social Research* (ICPSR) was established at the University

of Michigan as a repository for the data produced by the Michigan election studies and by other scientific surveys conducted by University of Michigan and other researchers.  ICPSR is now a general archive that receives and holds all sorts of social data collections (e.g., from commercial and media surveys, government agencies that collect statistics, election studies in other nations, etc.) and makes the data available to social science researchers, teachers, and students throughout the U.S. and the world.  UMBC is a member of ICPSR and all of its faculty, staff, and students have immediate and free access to its data using its website.  (Everyone in the world has direct access to ANES data using its website.)

## 3.      *The APSA/ICPSR SETUPS Teaching Modules*

Since 1974 the American Political Science Association (APSA) and ICPSR have collaborated in creating *Supplementary Empirical Teaching Units in Political Science* (SETUPS) modules based on major surveys and other data collection activities by political science researchers.  In particular, SETUPS modules based on the 1972 ANES and on every ANES from 1984 onwards have been created.  I used the most recent SETUPS module in POLI 300 for many years.  In the mid-1990s a new SETUPS module was created based on the ANES Cumulative Data File for all Presidential elections from 1972 through 1992. I immediately adopted this new *SETUPS: AMERICAN VOTING BEHAVIOR IN PRESIDENTIAL ELECTION 1972-1992*  for POLI 300, so that students could do "longitudinal" as well as "cross-sectional" exercises.  For the same reason, I continued to use it through Fall 2002 even after SETUPS based on the 1996 and 2000 ANES became available.  But in due course I recognized that even the 1992 election was increasingly "ancient history" for most POLI 300 students, and in 2003 I therefore updated the 1972-1992 SETUPS data to include data taken from 1996 and 2000 SETUPS.  In the summer of 2005, I extended the time series by extracting data from the recently released full 2004 ANES, and in summer 2009 I likewise extracted data from the just released 2008 ANES.  This handout (and other distributed material) replaces the *SETUPS: 1972-1992* booklet that POLI 300 students previously bought from the Book Center.

## 4.      *The NES 1972-1992 SETUPS*

The following description of the NES 1972-1992 SETUPS data is taken from the *SETUPS: 1972-1992* booklet.

> The data for this instruction package come from the American National Election Studies (ANES).  These large national surveys are conducted every election year by the Center for Political Studies at the University of Michigan. Data from six presidential elections, 1972 through 1992, were drawn from the ANES Cumulative Data File and modified for classroom use to create a data set that allows students to analyze voting in presidential elections over a two-decade period. All of the respondents interviewed in each of the six election studies are included in this data set, but in order to simplify the data set, only some of the variables are included. The 70 variables that were selected represent most of the important variables. Users should keep in mind that this data set is a collection of six surveys, each with its own set of respondents. Most questions were asked in each of the six years, so that information is available for every respondent regardless of the year in which the respondent was interviewed (except, of course, that there always is some missing data in any year). A few variables are available only for some years, and the codebook notes this for those variables.

## Using the Codebook

The codebook . . . provides both a description of the data and information necessary for using the data. Every available variable is listed in the codebook, and the entry for each variable contains the information needed to use the variable in the data analysis. Below is a sample codebook entry, with each specific item of information identified by the description that points to it. An explanation of each item of information follows the sample entry.

Variable Number   Variable Label        Text of Question or Description of Variable

⇩                      ⇩                        ⇩

**V04   PRESIDENTIAL VOTE**   "Whom did you vote for in the Presidential election?" (*Responses are categorized by the party of the candidate that the respondent voted for*.)

| 1 | Democratic |
| 2 | Republican |
| 3 | Other |
| 9 | NA |

⇧        ⇧

Value Codes   Value Labels

## Explanation of Codebook Information

1. *Variable number*. Each variable in the data set has been assigned a unique number, which is preceded by the letter "V" (for variable). This use of variable numbers provides a simple shorthand way of referring to variables in the data set.
2. *Variable label*. Each variable has been given a unique label. If certain statistical packages are used for the data analysis, the tables generated will have the appropriate variable labels printed on them as a convenient aid. Because there are maximum allowable lengths for these variable labels, they often have an abbreviated form.
3. *ICPSR reference number*.   [Deleted from this handout]
4. *Text of question or description of variable*. An explanation of the meaning of each variable is provided by an approximate description of the question asked or a general description of the variable. Also, if the question was not asked in each of the six elections, the years for which it is not available are listed. Many of the variables are relatively straightforward and need little explanation, but some types of variables require more thorough explanation and this is provided below.

   (a) There are several feeling thermometer items (V25-V28), which asked the respondent to indicate his or her feeling toward a specific candidate by placing that person on a feeling thermometer that ranges from 100 to 0 degrees, where 50 degrees represents a neutral feeling, higher temperatures represent warmer feelings, and lower temperatures represent cooler feelings. Placement on the feeling thermometers have been collapsed into five categories for ease of analysis.

   (b) There are a number of issue-position scales, each of which has a range of categories that represent possible positions that people might take on a specific issue. For example, there is an issue-position scale on defense spending (V47), and the possible positions on the scale run from "greatly decrease spending" to "greatly increase spending." Respondents were asked to place themselves on this scale according to their feelings on the issue. Only the end points of the scale are defined; respondents who feel that they fall between the two extremes can place themselves on one of the middle points. All of the issue-position scales

have this basic structure. They originally were seven-point scales, but they have been collapsed down to five categories for ease of analysis by combining the two end points with their adjacent categories.

    (c) There are candidate and party placement scales that indicate how the respondents felt that the candidates and parties should be placed ideologically. For example, in each year respondents were asked where they thought the Democratic and Republican presidential candidates stood on the ideology scale. These two candidate-placement scales (V35, V36) have five possible categories, running from liberal to conservative.

5.    *Value codes and value labels*. The possible values for each variable are given in the codebook. Both the numeric codes and a brief explanation of what the codes refer to are provided. If certain statistical packages are used for the data analysis, the tables generated will have the value labels printed on them as a convenient aid. Because there are maximum allowable lengths for these labels, they often have an abbreviated form. In the sample codebook entry given above, a "1" indicates a vote for the Democratic candidate, a "2" indicates a vote for the Republican candidate, and a "3" indicates a vote for an independent or third party candidate. Additionally, a code of "9" is used for respondents who do not fit into any of these categories. For this last group of respondents we have only "missing data." Missing data occurs because: (a) the question does not apply to the respondent — e.g., people who did not vote were not asked which presidential candidate they voted for; (b) the respondent refused to give a response or had no opinion; or (c) the interviewer failed to obtain or record the information for some other reason. The label "NA" is attached to this category to indicate that the item is "not applicable" or that the information was "not ascertained."

## 5.    *Updating the NES 1972-1992 SETUPS Data and Codebook*

As noted above, I have updated the *SETUPS: 1972-1992* data file and Codebook to include data from the data files accompanying *SETUPS: 1996* and *SETUPS: 2000* and from the full 2004 and 2008 ANES, creating a new data file and a new Codebook.  The new data file and Codebook are identical to *SETUPS: 1972-1992* with respect to the 1972-1992 data, so any analysis conducted on the new data file with respect to the years 1972 through 1992 should produce results identical to those one would get using the *SETUPS: 1972-1992* data for all variables other than V39 (which has been recomputed on the basis of V37 and V38 as recoded here).

However, you will find that a fair number of variables included *SETUPS: 1972-1992* are "not available" for 1996, 2000, 2004 and/or 2008.  With respect to 1996 and 2000, this is because either (i) these variables were not included in the 1996 and/or 2000 SETUPS (though they probably were included in the full 1996 and/or 2000 ANES questionnaires) or (ii) the 1996 and/or 2000 SETUPS variables were coded in a way that could not be reconciled with the prior coding.  With respect to 2004 and 2008, this is either because (i) these variables were not included in the most recent ANES questionnaires or (ii) special coding problems arose.  However one important variable (V65 INCOME) is included despite such coding problems, so you should take special note of the descriptions for V65A, V65B, V65C, V65D, and V65E  in the following Codebook.  The Codebook also notes minor coding discrepancies pertaining to V10 (INTEREST IN ELECTION), V67 (RELIGION), and V68 (CHURCH ATTENDANCE).   The 2008 ANES used a larger than average sample size and for a number of questions this sample was randomly split in half; respondents in one subsample were asked the "classic" version of the question, while those in the other subsample was asked a "revised" or "experimental" version of the question.  The 1972-2008 data includes only responses to the "classic" version, and the other half of the responses show up as "missing data" (NA).