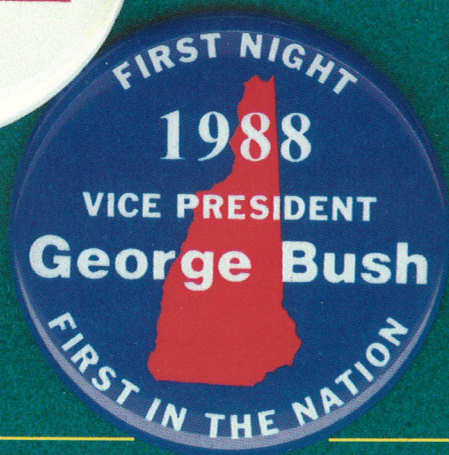
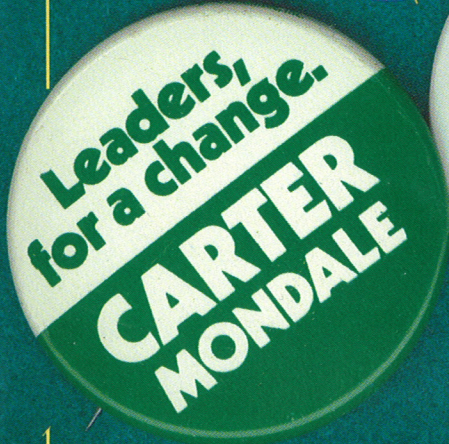
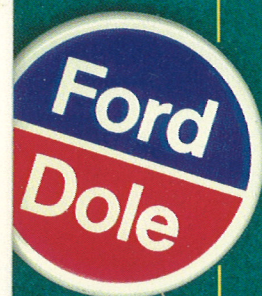
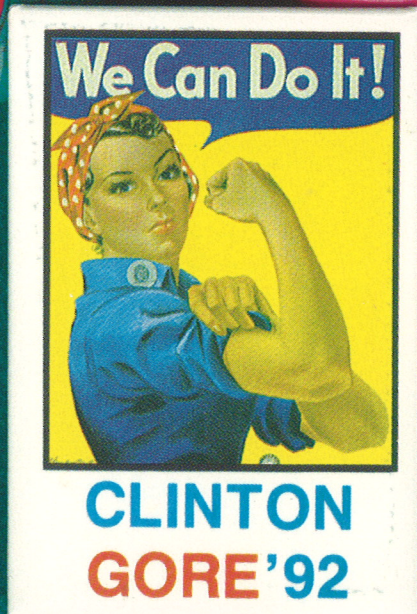
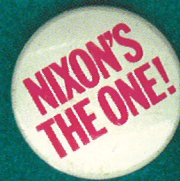


AMERICAN VOTING BEHAVIOR

IN PRESIDENTIAL ELECTIONS 1972 TO 1992



by Charles Drysby and Carmine Scavo

Supplementary
Empirical
Teaching
Units in
Political
Science

**AMERICAN VOTING BEHAVIOR IN PRESIDENTIAL ELECTIONS
1972 TO 1992**

Charles Prysby
University of North Carolina at Greensboro

Carmine Scavo
East Carolina University

1995

The American Political Science Association
Washington, D.C.

A SETUPS includes a monograph and a dataset. SETUPS data are distributed by the Inter-University Consortium for Political and Social Research through an agreement with the American Political Science Association.

SETUPS: American Politics

The SETUPS in this series are designed for use in courses on American government and politics. SETUPS (*Supplementary Empirical Teaching Units in Political Science*) were introduced in 1974, and they are being used in colleges and universities throughout the United States and in other countries. In the *APSA Survey of Political Science Departments for 1992–93*, 12% of the departments reported that SETUPS are adopted for computer based assignments in undergraduate courses.

Other SETUPS in American Politics currently available:

Presidential Popularity in America, Stephen Frantzich
Voting Behavior: The 1992 Election by Charles Prysby and Carmine Scavo
Financing Congressional Campaigns by Clyde Wilcox

American Voting Behavior in Presidential Elections from 1972 to 1992 is a new SETUPS that offers an accessible dataset drawn from the National Election Study Cumulative Data File. The dataset includes 70 variables, most of which were asked in each of the six years, and all of which are available for at least the majority of the years. Charles Prysby and Carmine Scavo who have developed SETUPS on the 1984, 1988, and 1992 presidential elections now provide faculty and students with an opportunity to examine voting behavior over a series of elections as well as for a single election.

© 1995 by the American Political Science Association, 1527 New Hampshire Avenue, N.W., Washington, D.C. 20036, under International and Pan-American Copyright Conventions. All rights reserved. No part of this book may be reproduced in any form or by any means without permission in writing from the publisher. Printed in the United States of America.

ISBN 1-878147-21-8

Table of Contents

Acknowledgments	ii
I. Presidential Elections, 1972-1992	1
From the 1930s to the 1960s	1
From the 1970s to the 1990s	2
Changes in the Electorate	5
II. Analyzing Voting Behavior	7
Sources of Individual Voting Behavior	7
Electoral Dynamics	9
Summary	10
III. Methods of Analysis	11
The Module Dataset	11
Survey Research Methods	12
Principles of Data Analysis	15
Computer-based Statistical Packages and Data Analysis	18
Conclusion	20
IV. Analysis Exercises	23
Exercise 1: Ideology and Partisan Choice	23
Exercise 2: Party Identification and Presidential Voting	25
Exercise 3: The Voting Behavior of Southern Whites	26
Exercise 4: The Gender Gap	27
Testing Your Own Theory	28
V. User Information	31
Using the Codebook	31
Explanation of Codebook Information	31
VI. Codebook	33
A. Voting Behavior and Related Items	33
B. Political Involvement Items	34
C. Candidate Image Items	34
D. Government Performance Items	35
E. Ideological Items	36
F. Attitudes on Specific Issues	37
G. General Orientations to Government	38
H. Social and Demographic Characteristics	38
References	41
Appendix: Technical Note	43

Acknowledgments

We gratefully acknowledge our gratitude to the American Political Science Association and the Inter-University Consortium for Political and Social Research for their support of this project. The APSA has supported the SETUPS series since 1974, when the first SETUPS were created. We are pleased that the APSA decided to support this SETUPS, which differs from the other voting behavior SETUPS modules by including data from a series of elections, rather than just a single presidential contest. Special thanks goes to Sheilah Mann, APSA Director of Educational Programs, for her role in this project. Janet Vavra, ICPSR Technical Director, provided documentation and technical assistance in the creation of the machine-readable files. Jennifer Harris and Shanna Reece provided capable research assistance at various stages of this project. Richard Niemi, Anne Permaloff, and Herbert Weisberg each read the entire manuscript, made many useful suggestions that improved the final product, and did so on very short notice. We appreciate the help supplied by these individuals. The authors naturally bear full responsibility for the contents of this book and the accompanying dataset.

Chapter I

Presidential Elections, 1972-1992

Presidential elections during the 1970s and 1980s were very favorable to Republicans. Republican presidential candidates won four of the five contests held from 1972 through 1988, leading many political analysts to classify the Republicans as the majority party in presidential elections. This majority status did not extend to congressional elections, however, as the Democrats controlled both houses of Congress throughout most of this period. The result was that the country had divided government more often than not. Moreover, Democrats regained control of the White House in 1992, raising questions about whether the Republican presidential majority was as secure as many had assumed. This mixed pattern of national election outcomes was subject to varying interpretations, and whether the nation was undergoing a Republican realignment was a much debated question among political commentators.

From the 1930s to the 1960s

To fully understand American electoral politics from the 1970s to the present requires information about prior voting patterns. From the 1930s to the 1960s, the Democrats were the majority party in the nation, a status that they achieved in the New Deal Realignment that followed Franklin D. Roosevelt's presidential victory in 1932, at the height of the Great Depression. The New Deal realignment established the Democrats as the party that favored an active federal government, particularly to ameliorate economic problems, and the party that favored working Americans. Republicans were viewed as the party that was more favorable to big business and the wealthy, an image that contributed greatly to their minority status. Republicans did have some successes during this period. Dwight D. Eisenhower was elected to two presidential terms in the 1950s, and Republicans briefly captured control of the Congress in the late 1940s and again in the early 1950s. But the Democratic Party clearly was the dominant party overall.

Democratic domination of national elections was based on their New Deal electoral coalition, composed of blue collar workers, ethnic minorities, and the South. During the 1930s, sharp class and ethnic divisions existed in American society (Ladd and Hadley 1978, 68-81). Working-class voters, especially those who belonged to unions, were considerably more Democratic than middle-class individuals. Many members of the working class were recent immigrants or first generation Americans, as the nation experienced a heavy wave of immigration in the early part of the twentieth century. Much of this immigration came from southern and eastern Europe, with a predominantly Catholic, Jewish, and Eastern Orthodox religious character. These new immigrants frequently felt discriminated against by more established ethnic groups, who generally were Protestants from the British Isles or Northern Europe. These working class and ethnic minority groups were concen-

trated in northern urban areas, giving rise to a clear urban-rural split in many northern states. Cities were Democratic strongholds while small towns and rural areas were heavily Republican.

The South also was an important component of the Democratic New Deal coalition. The sources of Democratic strength in the South were quite different from those in the North. The South in the 1930s was heavily Protestant and not very urbanized and industrialized, so Democrats could not win elections by drawing on the votes of groups such as Catholics, Jews, and unionized industrial workers. Instead, they drew support primarily from white Protestants who had already been solidly Democratic before the realignment of the 1930s. Democratic domination of the South occurred decades earlier and was a result of the Civil War and Reconstruction. Following the end of Reconstruction, white southerners began to support the Democratic Party as a means of asserting white supremacy, and by the early twentieth century every one of the 11 states of the old Confederacy was thoroughly dominated by the Democratic Party (Key 1949).

In the 1960s, the Democratic New Deal coalition began to fracture. The civil rights revolution radically altered the South. It enfranchised many southern blacks, who voted heavily for Democrats, but it also caused many conservative southern whites to abandon the national Democratic Party in presidential elections because of racial issues (Lamis 1984). In the North, the class divisions of the 1930s and 1940s began to diminish by the 1960s, as many of the economic issues of the Depression period were settled and an era of prosperity followed the end of World War II. Ethnic and religious divisions also diminished as the immigrant and first generation Americans were replaced by second and third generation Americans. The result was that northern working class and ethnic minority groups, while still more Democratic than the country as a whole, were much less distinctively so by the end of the 1960s (Ladd and Hadley 1978, 235-237).

Developments in the late 1960s were crucial to the resurgence of the Republican Party in presidential elections. In 1968 the nation was deeply involved in the Vietnam War. Questions increasingly were being raised about whether we should have intervened, why victory was so elusive, and what should be done to end our entanglement. Protests against the war combined with civil unrest related to racial issues, including several major riots in urban black ghettos, to create a situation of domestic turmoil. Incumbent President Lyndon Johnson, who had won a landslide victory in 1964, faced intense antiwar opposition from fellow Democrats. Confronted with the internal divisions in his own party, Johnson decided in the spring of 1968 not to seek reelection. In a tumultuous convention that summer, a deeply divided Democratic Party nominated Vice President

2 Presidential Elections

TABLE I-1: PRESIDENTIAL ELECTION OUTCOMES, 1972–1992

<i>Candidate (Party)</i>	<i>Percentage of Popular Vote Total</i>	<i>Electoral College Vote</i>
1972		
Richard Nixon (Republican)	60.7	520
George McGovern (Democrat)	37.5	17
1976		
Jimmy Carter (Democrat)	50.1	297
Gerald Ford (Republican)	48.0	241
1980		
Ronald Reagan (Republican)	50.8	489
Jimmy Carter (Democrat)	41.0	49
John Anderson (independent)	6.6	0
1984		
Ronald Reagan (Republican)	58.8	525
Walter Mondale (Democrat)	40.6	13
1988		
George Bush (Republican)	53.4	426
Michael Dukakis (Democrat)	45.6	112
1992		
Bill Clinton (Democrat)	43.3	370
George Bush (Republican)	37.7	168
Ross Perot (independent)	19.0	0

Source: Various editions of *America Votes* (Washington, DC: CQ Press).

Hubert Humphrey. The Republicans chose Richard Nixon, who had served as Eisenhower's Vice President. A third candidate, Alabama Governor George Wallace, ran as an American Independent Party candidate, stressing a very conservative message in both foreign and domestic matters. Nixon won a very close presidential election with 43% of the vote, beginning a period in which Republicans had the advantage in presidential contests.

From the 1970s to the 1990s

Nixon's election in 1968 was seen by some as the beginning of a Republican era. In terms of presidential elections, this has been the case for the past two decades. Republicans won four of the six presidential elections held from 1972 to 1992. But attributing these victories simply to a Republican advantage in presidential elections during this time period obscures how unpredictable these elections were. Few of the outcomes would have been predicted by most analysts three or four years prior to the election. Each election outcome depended on a set of factors unique to that election. The important aspects of each election are summarized below, and Table I-1 contains the popular vote and electoral college outcomes for each race.

1972

Although Nixon was narrowly elected with only 43% of the vote in the three-way presidential race in 1968, he was

easily reelected in 1972. The Democratic Party, still deeply fractured ideologically, nominated Senator George McGovern, who was perceived by many voters as too far to the left (Miller et al. 1976). McGovern was seen as favoring a rapid troop withdrawal from Vietnam, a substantial reduction in defense spending, and a significant expansion of domestic social welfare programs. The Vietnam War was still a major issue in 1972, but U.S. involvement was much lower than in 1968, as the Nixon administration had been withdrawing American troops and turning more of the fighting over to the South Vietnamese forces. While there still was considerable polarization over this issue, the majority of voters seemed to approve of Nixon's policy (Miller et al. 1976).

Voters also questioned McGovern's judgement and leadership ability. One episode which contributed to voter uneasiness about McGovern's presidential character concerned the selection of his running mate. McGovern's first choice for Vice President, Missouri Senator Thomas Eagleton, withdrew after information about his psychiatric treatment for depression became public knowledge, but only after McGovern displayed considerable indecision on the matter. Although Nixon had some liabilities when it came to personal characteristics (the nickname "tricky Dick" had long been applied to him), most voters felt that he was experienced and a capable leader.

The landslide victory for Nixon gave rise to predictions that the Republicans were about to replace the Democrats as the majority party in American politics. But Republican hopes for rapid electoral gains were dashed by the Watergate scandal. Less than two years after his reelection, a disgraced Nixon resigned from office and was replaced by Gerald Ford, who had only recently replaced Spiro Agnew as Vice President. The nation suffered one of its most serious political crises in modern times. Republicans approached the 1976 presidential election far more vulnerable than they had appeared four years earlier, while the Democrats appeared rejuvenated by these recent developments.

1976

The Republicans selected Gerald Ford as their presidential candidate in 1976, but only after a bitter nomination contest between Ford and former California governor Ronald Reagan, who appealed to the more conservative elements in the party. The fact that the incumbent president (even though he was not an elected incumbent) was seriously challenged for the nomination of his party showed that the Republicans were deeply divided, a fact that may have cost Ford the general election. The Democrats, by contrast, were well unified behind their nominee, Jimmy Carter, a little-known former Georgia governor.

The general election was a closely fought contest between these two candidates, with considerable attention paid to their personal characteristics. Ford stressed his record of federal governmental service, which contrasted sharply with Carter's lack of experience in Washington. Republicans also accused Carter of being vague and ambiguous on the issues, a charge that had been made by some Democrats during the primaries. Carter emphasized the fact

that he was an “outsider” who could not be blamed for the mess in Washington, referring in part to Watergate. Although Ford was not personally involved in Watergate, he did issue a general pardon to Nixon for any possible wrongdoing in Watergate. That action, along with simply being a Republican, linked Ford at least indirectly to Watergate.

The economy was a key factor in the election (Pomper et al. 1977, 36-43). Carter criticized the Ford administration for allowing unemployment to rise. The Democrats argued that the federal government needed to insure full employment so that all who wanted work could find it. Republicans countered that the Democrats were not sufficiently concerned with fighting inflation. Also, energy policy became an important economic issue in 1976, as the nation recently had to cope with oil shortages resulting from a boycott imposed by oil exporting countries. Ford emphasized the need for low energy prices in order to have economic growth. His solution was to rely on private enterprise and to avoid environmental restrictions that might jeopardize the goal of abundant and cheap energy. The Democrats favored more governmental action in this area, arguing that protecting the environment should not be sacrificed and that the nation could not simply rely on the actions of private enterprise to ensure that energy supplies would be sufficient. The two candidates also disagreed on tax policy, with Carter claiming that upper income groups were not paying their fair share, while Ford proposed cutting taxes for the middle class.

While economic problems received the most attention in the 1976 presidential campaign, the candidates did discuss other issues as well (Pomper et al. 1977, 43-53). Abortion divided the candidates somewhat, with Ford favoring a constitutional amendment that would allow states to enact restrictions on abortion and Carter opposing such an amendment, although he claimed that he was personally opposed to abortion. Carter also criticized the foreign policy of the Ford administration. He claimed that the United States had sacrificed morality in its foreign policy by supporting dictators in several countries, had failed to be firm in dealing with the Soviet Union, and had been too secretive in the making of foreign policy. Republicans not only defended the foreign policy of the Ford administration; they also attacked Carter for proposing cuts in the defense budget that they claimed would threaten national security.

Although they differed on a number of issues, both Carter and Ford were considered moderates. Carter generally was perceived as somewhat liberal and Ford as somewhat conservative, but few voters thought that the differences between the two were extreme, as was the case in 1972, for example. In the end, Carter won a narrow victory, perhaps in part because more voters considered themselves Democrats in 1976. Ford became the first incumbent president since Herbert Hoover in 1932 to lose a reelection bid. He was not the last, however.

1980

Four years later Carter lost his reelection bid. Burdened with an economy suffering from both high unemployment and high inflation, Carter was seriously challenged for the

Democratic nomination by Senator Ted Kennedy, who represented a more liberal wing of the party. As was the case with Ford in 1976, although Carter was able to win renomination, the attacks from within his own party weakened his reelection prospects. The Republicans nominated Ronald Reagan, who had run a strong race against President Ford for the Republican nomination four years earlier.

Reagan campaigned on a clearly conservative platform, emphasizing the need to cut taxes, cut domestic spending, increase defense spending, and balance the budget. Equally important, Reagan attacked the failures of the Carter administration, claiming that the economy was a disaster and that America was losing its military strength (Pomper et al. 1981, 38-61). Contributing to public dissatisfaction with the foreign policy of the Carter administration was the fact that Iranian militants had taken control of the American embassy in Tehran and had been holding the American diplomats hostage since November, 1979. The inability of Carter to achieve their release led to an image of a militarily inferior third world nation being able to dictate to the United States.

Dissatisfaction with the performance of the Carter administration, especially concerning the economy, rose to the point that Carter's approval rating in the Gallup Poll slipped below 30% in mid-1980 (Lewis-Beck and Rice 1992, 40). The election probably would have been a more lopsided victory for Reagan, but polls showed that the electorate had misgivings about the Republican nominee as well (Pomper et al. 1981, 105). Reagan was seen by many as inexperienced in national politics and prone to impulsive judgments. He also was considered by many as too conservative (Abramson, Aldrich, and Rohde 1982, 132-138).

For most of the campaign the outcome appeared uncertain, but there was a significant shift toward Reagan in the final week of the campaign, and he won by a comfortable margin. For those who were disenchanted with both major party candidates, John Anderson, running as an independent, offered a third choice; he ran well in the early polls but in the end won only about six percent of the vote. Reagan's presidential victory came with substantial coattails for Republican congressional candidates. After the election, Republicans held a majority of the U.S. Senate seats, for the first time since 1954. While Democrats retained their majority in the U.S. House, Republicans made substantial gains, and the combination of Republicans and conservative southern Democrats constituted a working majority.

1984

The 1980 election again gave rise to speculation about an emerging Republican majority. At the presidential level, it began to appear that the Republicans were becoming the dominant party. In 1984, Reagan easily won reelection over the Democratic nominee, Walter Mondale, who had been Vice President under Carter. The Reagan campaign aimed at making the election a referendum on the performance of the Reagan administration and on the personal qualities of the President. Given the peace and prosperity in the nation and the warm feeling that most

4 Presidential Elections

voters had toward Reagan as an individual, these appeared to be sizable advantages.

Reagan had been elected in 1980 in large part because of the economic problems facing the country. By 1984 the economy was much healthier. Unemployment and inflation were down. The country was in the midst of a substantial economic expansion that followed the recession of 1982. Moreover, the United States was not involved in any wars, did not have troops engaged in hostilities, and did not face any other foreign crisis (such as the Iran hostage situation of 1980). Thus, Reagan was able to run on a platform of peace and prosperity (Pomper et al. 1985, 35-57).

Mondale claimed that all was not well with the nation. He criticized the Reagan administration for greatly increasing the budget deficit, due to the tax cuts and spending priorities implemented during Reagan's first term, and claimed that this would be detrimental to the long-run economic health of the nation (Pomper et al. 1985, 43-44). Taxes would have to be raised, according to Mondale. Reagan denied that a tax increase was necessary, arguing that we would outgrow the budget deficit. Mondale also charged that the tax cuts and spending changes enacted during the first Reagan administration were too favorable to business and upper income groups. While these charges may have resonated well among some voters, it was not enough to compensate for the advantages that Reagan had in other areas.

Mondale and Reagan presented the voters with a clear choice between a liberal and a conservative. On almost every major issue—abortion, defense spending, civil rights, welfare programs, environmental protection, taxation—the candidates divided in an ideologically predictable manner (Pomper et al. 1985, 35-57). But the clear candidate differences did not mean that issues played a major role in the outcome (Abramson, Aldrich, and Rohde 1987, 172-183). Instead, the important factors appear to have been the personal appeal of Reagan and the satisfaction with the performance of his first term (Abramson, Aldrich, and Rohde 1987, 187-204). Reagan was seen by many voters as likeable, trustworthy, and a strong leader. He was especially effective in communicating a positive image on television. Mondale, on the other hand, was less inspiring, a wooden speaker, and not particularly comfortable on television, although he did do well in one of the debates with Reagan (Pomper et al. 1985, 76).

Reagan won a landslide victory in 1984, both in the popular and the electoral college vote. But his was a personal victory, not a party victory. Republicans held fewer seats in the House after the 1984 election than they had after the 1980 election, and in the Senate they had only the same number. Two years later, Republicans lost their Senate majority in the midterm elections. Many analysts concluded that the Republican advantage was limited to the presidential level.

1988

Four years later, Reagan's Vice President, George Bush, was elected, giving the Republicans three consecutive presidential victories, something that neither party had achieved

since the 1940s. The Bush campaign touted the accomplishments of the Reagan administration. The relatively sound state of the economy, along with the peaceful international situation, gave Republicans an advantageous situation for the election. Although Bush lacked the personal appeal of Reagan, he was experienced in national and world affairs. Positive themes were blended with negative ones in the Bush campaign. Republicans aimed at characterizing the Democratic nominee, Massachusetts Governor Michael Dukakis, as an extreme liberal who did not share the values of ordinary Americans (Pomper et al. 1989, 81-83). Dukakis's actions as governor were scrutinized, and he was attacked for several actions, including some hot symbolic issues. Special attention was focused on the Massachusetts prison furlough program, especially the infamous Willie Horton case (Horton was a black prisoner who raped a woman while on a weekend furlough). Criticism also was leveled at Dukakis's veto of a bill passed by the state legislature that would have mandated daily recitation of the pledge of allegiance in the public schools.

While the Republicans ran a skillful campaign, Dukakis ran what was generally considered to be a fairly inept campaign. The Republican attacks went unanswered for a long time. The Democratic themes were not presented in a clear and consistent fashion. Finally, Dukakis did not distinguish himself in the televised debates with Bush. Although there were many policy differences between the two candidates, including health care, taxation, abortion, and defense spending, the election was remarkably devoid of discussion of important public policy issues. Bush won a comfortable victory, but again Republicans had to be satisfied with capturing the White House. They actually lost a few seats in Congress, leaving the Democrats with comfortable majorities in both houses.

1992

This string of three Republican presidential victories, combined with high approval ratings for Bush in his first two years as president, led most political analysts to predict reelection for Bush and possible gains for Republicans in congressional and state elections. But once again American electoral politics proved difficult to predict. The declining health of the national economy produced a precipitous drop in Bush's approval ratings, opening the way for a Democratic victory. One sign of Bush's vulnerability was the challenge that he received for the Republican nomination from Pat Buchanan, a conservative commentator and former member of the Nixon and Reagan administrations. Buchanan did well in the early primaries, receiving nearly one-third of the votes, largely by attacking Bush for failing to adhering to conservative principles (Pomper et al. 1993, 46-50).

Although many of the strongest potential Democratic candidates had opted not to run, making their decision at a time when it appeared that Bush would be difficult to defeat, a number of candidates did compete for the Democratic nomination, which was won by Arkansas Governor Bill Clinton. Although questions were raised during the primary campaign about alleged extramarital affairs and avoid-

ing the military draft in the 1960s, Clinton nevertheless managed to win enough primaries to become the Democratic presidential nominee. Still, damaging questions had been raised about his character. Also running against Bush was Ross Perot, a wealthy Texas businessman who spent millions of his own money in what was one of the most interesting and unusual independent presidential candidacies in modern political history.

Both Clinton and Perot criticized Bush's handling of the economy, putting the President on the campaign defensive. Clinton also emphasized the need for change, including improving the nation's health care system, reforming the welfare system, and making the country more economically competitive (White 1993). Perot stressed the need to bring the deficit down and argued that he was an outsider who would not conduct "business as usual" in Washington (Feigert 1993). Both Clinton and Perot argued that Bush was an ineffective leader.

The Bush campaign concentrated on raising questions about Clinton's presidential character, including charges that he was evasive on the issues, and on portraying Clinton as a "tax and spend" liberal (Pomper et al. 1993, 74-80). However, Bush's credibility on the tax issue was undercut by the fact that he supported a tax increase in 1990 after pledging in 1988 never to raise taxes. Although Bush acknowledged that economic problems existed, he argued that the media was exaggerating the situation and that much of the blame should be placed on the Democratic Congress.

Clinton attempted to avoid the tag of being too liberal by claiming that he was a new Democrat, not tied to the failed policies of the past (White 1993). On some issues, Clinton took a conservative position, favoring capital punishment and proposing that welfare recipients be limited to two years of benefits. Still, on most issues—including abortion, gun control, civil rights, defense spending, environmental regulations, and health care—the two presidential tickets were ideological distinct.

The final result was a three-way race that Clinton won with only 43% of the vote. Nearly one voter in five opted to vote for Perot. Once again an incumbent president failed to be reelected, for the third time since 1976. The 1992 election outcome complicated any attempt to give a simple interpretation of how the electorate was shifting. The idea that Republicans were solidifying their grasp on the White House had to be questioned after 1992. Also, Democrats retained control of the U.S. House throughout the Reagan and Bush presidencies and recaptured the U.S. Senate in 1986, leading some observers to argue that Republican gains were not very deep. Indeed, one common interpre-

tation of the 1980s was that the electorate was moving toward the Republicans for the presidency but retaining an affinity for Democrats in Congress (Abramson, Aldrich, and Rohde 1991, 285-290). But even congressional elections seemed to display unexpected patterns. While Clinton was winning the presidential race in 1992, Democrats were losing seats in the U.S. House, a historically unusual situation, but one that duplicated the 1988 pattern. And in 1994 the Republicans gained an extraordinary number of seats in the midterm congressional elections, capturing a solid majority of seats in both houses of Congress for the first time since 1952 (Abramson, Aldrich, and Rohde 1995, 317-350). In sum, recent patterns suggest an inconsistent and unpredictable electorate.

Changes in the Electorate

The recent volatility of the American electorate may reflect some fundamental changes in attitudes and orientations over the past few decades. Beginning in the 1960s, voters have become increasingly independent in their partisan orientations. Although most voters still are willing to express some affinity to one of the two major parties, few call themselves strong partisans, a significant shift from earlier patterns (Abramson, Aldrich, and Rohde 1995, 228-231). This growing lack of partisan attachments within the electorate has accompanied a rise in candidate-centered election campaigns. Changes in partisan orientations have consequences for behavior. Voters are now more willing to split their ballot in a given year, such as voting for a Republican presidential candidate but a Democratic congressional candidate. Voters are also more likely to switch the partisan direction of their vote for a given office across years, voting Republican in one presidential election and Democratic in the next, for example.

The contemporary electorate also is a cynical, suspicious, and distrustful set of voters. This also marks an enormous change from the 1960s. In 1964, nearly 80% of the citizenry felt that they could trust the government in Washington to "do what is right" at least most of the time; in 1992, only about 30% would make that claim (Luttbeg and Gant 1995, 137). This decline in trust in government began in the late 1960s, accelerated during the 1970s, and has remained at a relatively unfavorable level throughout the 1980s. Voters also display lower levels of political efficacy. As compared to the 1960s, many more voters believe that they do not have much influence on what government does and that public officials do not care what most people think (Luttbeg and Gant 1995, 139). These attitudinal changes may at least partly explain the willingness of recent electorates to turn incumbent presidents out of office.

Chapter II

Analyzing Voting Behavior

Analyzing the behavior of voters in recent presidential elections requires a theoretical understanding of the basic forces that influence electoral behavior. Two major concerns characterize the research in this area. One concern is with explaining the results of particular elections by identifying the sources of individual voting behavior. We attempt to understand an election outcome by analyzing how and why the voters made up their minds. A second major concern in voting research focuses on the dynamics of electoral behavior. In this case, the emphasis is on the nature and sources of changes in voting patterns over time. These two concerns are complementary, not contradictory, but they do emphasize different sets of theoretical questions. For our purposes, these two concerns provide a useful basis for discussing key aspects of voting behavior.

Sources of Individual Voting Behavior

On what basis do voters decide how they will cast their ballot? Several basic factors can be identified as reasons for preferring one candidate to the others in an election. The voter may decide on the basis of one or more of the following considerations: (a) evaluations of the personal qualities of the candidates; (b) general assessments of the performance of government; (c) orientations on specific issues of public policy; and (d) basic loyalty to or preference for a particular political party. Candidate qualities and government performance are distinctly short-run forces, capable of substantial shifts from one election to the next. Party loyalties are much more stable in the short run. Issue orientations fall somewhere in between. While the specific issues crucial in presidential elections can change dramatically, many basic policy questions (e.g., defense spending, welfare programs, abortion) stretch across several elections, with partisan differences on these issues remaining relatively constant.

Voters frequently judge presidential candidates by their personal characteristics. They form images of the personal qualities and abilities of the candidates, and these perceptions are important influences on the vote. Included among the relevant characteristics are such things as the knowledge, experience, honesty, morality, compassion, competence, and leadership ability of the candidates (Page 1978). Such considerations seem to have played an important role in many recent presidential campaigns. In 1992, Bill Clinton was accused of marital infidelity and of avoiding the military draft during the Vietnam War, while President Bush was attacked for being a weak leader. Similarly, Michael Dukakis was criticized by Bush in 1988 for being unpatriotic and lacking American values, while questions were raised about Bush's leadership qualities by the Democrats. While some candidates have been hurt by negative perceptions of their personal characteristics, others have

benefited from very positive ones. For example, much has been made of Ronald Reagan's strong personal appeal to voters.

There is substantial variation from one election to the next in voter perceptions of the personal characteristics of the presidential candidates. The Republican candidate may have a significant advantage when it comes to personal characteristics in one year, while the opposite may be true in the following election. The uncertainty in knowing very far ahead of the election how the presidential candidates will be perceived in personal terms is one factor that makes predicting the outcome of presidential elections a difficult task. A number of items in the dataset for this module deal with respondent evaluations of the personal qualities of the presidential candidates, allowing us to examine the influence of these factors in particular elections and to compare this influence across elections.

The role of public policy issues in elections is of particular interest to political analysts. The term issue sometimes is used loosely to refer to anything that is a source of conflict or contention, but that is not its meaning here. We are referring to public policy issues, meaning questions of what the government should or should not do. Policy issues involve conflict over the direction of government action. Some policy issues in an election may be quite specific, such as the conditions under which abortion should be legal. Other policy issues are general, dealing with broad approaches to problems, such as whether the federal government should cut spending on social welfare programs. Elections are widely justified as providing a means for citizens to influence governmental decisions by choosing among contenders for office, and the assumption often is that the electorate will shape government policy by selecting candidates on the basis of policy issues. When this does not appear to be the case, political commentators often are quite critical, and we frequently hear complaints that the candidates in an election are failing to clearly address the issues.

For a policy issue to affect the vote decision, voters must have opinions on the issue and must perceive differences between the candidates on the issue. Even on important issues, some voters will fail to meet these conditions. Some will have opinions that are too weak and unstable to provide a basis for evaluating the candidates, while others will fail to see any significant differences between the candidates on the issue (Campbell et al. 1960, 167-187). But many voters will have definite opinions and clear perceptions of candidate differences on at least some issues, and this will be true especially when the candidates clearly articulate their differences (Nie, Verba, and Petrocik 1976, 156-173).

Specific issues have received considerable discussion in many recent presidential campaigns. Health care, for ex-

8 Analyzing Voting Behavior

ample, was a key issue in 1992, as Clinton called for a national health care plan. Defense spending has been important in some campaigns, especially 1980, when Reagan argued that substantial increases were needed in defense spending to restore American military strength. Abortion has divided the presidential candidates in a number of recent elections. Taxes and government spending almost always receive considerable attention in presidential campaigns. The data for this module contain measures of how respondents felt about a number of very basic policy issues, allowing us to examine the impact of these attitudes on the vote in recent presidential elections.

Although the specific policy issues of greatest importance vary considerably from one election to the next, the general pattern has been for the Democratic candidate to be perceived as more liberal overall and the Republican candidate more conservative overall. The extent of the differences between the candidates has varied. In 1976, the two candidates were seen in fairly moderate terms, with the Democrat slightly left of center and the Republican slightly right of center (Abramson, Aldrich, and Rohde 1982, 126-129). In other years, such as 1984, the ideological differences were much greater (Abramson, Aldrich, and Rohde 1987, 168-172).

Rather than choosing among candidates on the basis of specific issues of public policy, voters may rely more on general evaluations of the performance of government. A presidential election is, at least in part, a referendum on the performance of the incumbent administration. This is especially true when the incumbent president is running for reelection, which was the case in all but one of the elections held between 1972 and 1992. In the one exception, 1988, the incumbent Vice President was the presidential candidate of his party. Retrospective evaluations of government performance are an important determinant of voting behavior, and this effect should be distinguished from the influence of policy issues (Fiorina 1981). Policy issues involve differences over what the government should do; they are prospective in nature. Performance evaluations involve differences over how well the government has done; they are retrospective in nature. Quite often, we find agreement over what the government should accomplish, but disagreement over how well the goals have been achieved. Basic goals such as low unemployment, low inflation, steady economic growth, national security, and world peace are shared by all. Candidates do differ in their prescriptions for economic health or national security, but discussions of the details of macroeconomic theory, or of defense strategies, may not be scrutinized by many voters. More relevant are general perceptions of whether the economy or national security has improved or declined.

The importance of these factors is reflected by the emphasis given to them in recent presidential elections. The condition of the economy was a critical factor in the 1992 presidential election. Bush's opponents claimed that he was responsible for the poor health of the economy, while the President claimed that the nation's economic problems were being exaggerated by the media. Republicans in 1980 sought to tie negative evaluations of the economy and the

international environment to perceptions of President Carter's competence. In 1984, the Republicans emphasized the improvement in the economy and the international environment that occurred during the Reagan administration. Similar claims were made by Republicans in 1988, while Democrats countered that everything was not so well off.

Although evaluations of government performance can involve more than just the condition of the economy, it appears that economic conditions are particularly important in influencing the performance evaluations of voters (Kiewiet 1983). Moreover, economic conditions tend to shift dramatically. In some years, such as 1980 or 1992, the economy was badly off, especially in the eyes of the public, whereas in other years, such as 1972 or 1984, there was considerable satisfaction with the state of the economy. Still other years, such as 1976 or 1988 received more mixed evaluations from the public. This dataset includes several measures of perceptions of government performance and related factors, including items dealing with economic conditions, thereby allowing for an examination of the role of retrospective evaluations in presidential elections.

Finally, a voter may cast a ballot for a candidate because of the candidate's party affiliation. Most voters express some identification with one of the two major political parties, although these attachments tend to be substantially weaker than in the past. These basic partisan loyalties may influence their votes. Party identification normally is measured by asking individuals whether they consider themselves to be a Democrat, Republican, or independent. Those indicating Democratic or Republican can then be asked whether they are a strong or a weak Democrat or Republican, while those claiming to be an independent can be asked whether they feel closer to one of the two political parties. This yields a sevenfold classification: strong Democrat, weak Democrat, independents closer to the Democrats, independents not closer to either party, independents closer to the Republicans, weak Republicans, and strong Republicans.

The direct influence of party identification on the vote probably is small in presidential elections. By this we mean that there probably were few voters in recent elections who cast their ballots for or against a presidential candidate simply because he was a Democrat or a Republican. But the indirect influence of party identification is much greater, in that partisan loyalties influence evaluations of candidates, assessments of government performance, and perceptions of political events. Put simply, party identification is a perceptual screen—a pair of partisan tinted eyeglasses through which the voter views the political world. Party identification is less important now than in the past, but it is still a significant factor for explaining political orientations and behavior. In fact, party identification is still a very good predictor of how individuals vote.

Evaluations of candidate personal characteristics, assessments of government performance, and orientations on public policy issues are not only influenced by party identification. They affect each other. Strong conservatives are not only more likely to agree with the issue positions of a conservative candidate; they also will be likely to have a favorable view of the personal characteristics of that

candidate. Voters who feel that the incumbent President has not managed the economy very well are likely to unfavorably judge many of the President's personal qualities, especially those regarding leadership. Although the three attitudes discussed above are conceptually distinct, they are empirically interrelated. In sum, there are multiple reasons why people vote the way that they do. The relative weight of these factors may vary among individuals. Some people may vote more on the basis of the personal characteristics of the candidates, others more on the basis of issues of public policy, and still others on the basis of retrospective performance evaluations. A few may cast a ballot largely on the basis of party identification. The importance of the above factors also can vary from one election to another. While it is easy to outline the possible factors that motivate the vote, it is more difficult to specify exactly how important each factor was in determining the election outcome.

Some of the important controversies in the literature on voting behavior involve the relative importance of different factors in determining individual voting. During the 1950s, considerable emphasis was placed on the role of party identification as a determining factor (Campbell et al. 1960). Voters were viewed as politically unsophisticated, responding primarily to basic partisan loyalties and lacking in an understanding of the issues underlying the election. During the 1960s and 1970s, the literature began to focus more on the role of issues in elections (Key 1966; Nie, Verba, and Petrocik 1976). Voters were now seen as capable of choosing among candidates on the basis of issues. In the 1980s, considerable attention was paid to retrospective voting, especially in response to economic conditions (Fiorina 1981; Kiewiet 1983). To some extent, these changes in emphasis reflected changes in voting behavior, but they also represented changes in the theories and methodologies used to study voting behavior.

Controversies in explaining voting behavior have encompassed one of the most basic concepts in the field, party identification. Earlier literature stressed that party identification was learned at a relatively early age and remained relatively stable throughout the life span (Campbell et al., 1960). Once formed, party identification exercised considerable effect on an individual's political attitudes, but there was little influence in the opposite direction. Recently there has been more acceptance of an alternative view of party identification. This conception sees party identification as reflecting as well as influencing individual attitudes. Voters can and do shift their party identifications in response to their evaluations of the parties in terms of issues and performance. Even this revised view conceives of party identification as a fairly stable orientation—more stable than issue orientations or performance appraisals, for example—but it gives more consideration to how individuals might bring their party identification into line with other attitudes.

Analyzing the factors that affect the vote decision also leads to a consideration of how voters form their perceptions and evaluations. Why, for example, do some voters feel that a candidate is trustworthy, while others have the opposite opinion? What leads voters to have favorable or unfavorable assessments of government economic perform-

ance? As suggested above, party identification plays a role, but it does not fully explain the formation of these attitudes and orientations. Members of the same party often will have divergent views. While we do not have the space here for an extended discussion of this topic, it is clear that personal experiences, basic values, membership in various groups, exposure to the mass media, and political discussions with friends all play an influential role. One set of factors, social and demographic characteristics, are frequently examined, in part because they form basic cleavages in society. Race, religion, region, and socio-economic status have divided voters for decades. Other characteristics, such as gender or age, have played a role in recent elections as well. The influence of these social and demographic characteristics on attitudes and voting is due to the fact that they shape the experiences, values, interests, and information sources to which the voters are exposed.

Electoral Dynamics

Election results often change dramatically. A lopsided victory for one party may be followed by a landslide for the other party in the following election, as the discussion in Chapter I shows. Electoral changes can be divided into two types: short-run and long-term. Short-run changes are the result of fluctuations in factors that are specific to an election, such as the characteristics of the candidates or the particular issues that are salient. These short-term factors may be moderately favorable to the Democrats in one election, strongly favorable to the Republicans in another, and evenly divided in a third.

Voter perceptions of candidate personal characteristics and evaluations of government performance are primarily short-run factors. They tend to be specific to a particular election, and they easily change from one election to the next, as the candidates and events change. The fact that one party has a big advantage on these two factors in one election does not mean that it will have a similar advantage in the following election. However, these perceptions and evaluations can have long-term effects. Voters may develop an identification with a party because of favorable attitudes toward the candidates, policies, and accomplishments of the party—or because of unfavorable attitudes toward the opposing party. A voter's party identification may represent, at least in part, a summary evaluation of how the voter has perceived recent political history (Fiorina 1981).

Long-term shifts result from alterations in basic loyalties and represent changes that last beyond a particular election. The most significant long-term change occurs when there is a critical realignment of the party system, which refers to a relatively rapid, fundamental, and durable alteration in the pattern of party loyalties (Sundquist 1983). Realignments occur infrequently; the last major upheaval of the party system occurred in the 1930s, and before that in the 1890s and 1850s. Of course, in any time period there is some change in party loyalties, but only rarely is it substantial enough to qualify as a realignment.

The 1930s realignment shaped the current party system. The Great Depression acted as the catalyst for a transformation of the party system that moved the Democrats from

10 Analyzing Voting Behavior

minority to majority status at the national level. The New Deal Democratic coalition that put Franklin D. Roosevelt in the White House and the Democratic Party in control of Congress combined support from the working class and various ethnic and minority groups with already existing strength in the South. The basis of Democratic appeal to blue-collar workers, low-income individuals, and recent immigrant groups (largely Catholics and Jews from southern and eastern Europe) was the party's liberalism in economic matters. Roosevelt and the Democrats favored federal government activity to combat the Depression and proposed programs to benefit disadvantaged groups. The Republicans, who appealed more to the middle-class, business groups, and northern white Protestants, were critical of this expansion of government interference in the economy and creation of a variety of social welfare programs. By the late 1930s, the lines between the two parties were clearly drawn, both in ideological and socioeconomic terms (Ladd and Hadley 1978).

Although the impact of the New Deal realignment has remained for decades, it has done so only in a diluted and revised form. The class cleavages that were so clear in the 1930s and 1940s have diminished greatly in recent years (Abramson, Aldrich, and Rohde 1995, 146). Differences between Protestants and Catholics in their partisanship have eroded somewhat as well (Abramson, Aldrich, and Rohde 1995, 146). And the South, once solidly Democratic, has voted disproportionately Republican in recent elections (Black and Black 1992, 58-76). Still, the party images of an earlier era persist. Democrats remain thought of as the party that favors bigger government, more spending on domestic programs, and helping those at the bottom of the economic pyramid. Republicans continue to be perceived as favoring limited government, less spending on domestic programs, and fewer restrictions on business enterprises. These are not baseless images. They reflect continuing fundamental differences between the parties.

At the same time that many of the old partisan differences have diminished, new divisions have emerged. Beginning in the 1960s, blacks began to vote in greater numbers and to cast their ballots overwhelmingly for Democratic candidates. They now are one of the most loyal components of the Democratic coalition (Abramson, Aldrich, and Rohde 1995, 144-147). This development may be one factor in the declining support that Democrats have

received from white voters (Edsall and Edsall 1991). In the 1980s, a gender gap emerged, with men more likely than women to vote Republican (Mueller 1988). Marital status also has become important in recent elections; married individuals now are more Republican in their voting than are single people (Weisberg 1987). A new religious split has been developing, with Republicans appealing more to a "New Christian Right," which emphasizes traditional moral values (Wald 1987). And much has been made of Republican tendencies among suburban voters, who constituted a majority of the electorate for the first time in 1992 (Schneider 1992).

While there has been a great deal of change in recent years, it has not added up to a fundamental realignment, on the order of what happened in the 1930s. The changes have occurred more slowly and have been less complete. In fact, some of what has occurred might be better described as dealignment: many voters have become less attached to either party, with the result being a very volatile electorate that is highly responsive to short-term forces (Wattenberg 1994). The diminished loyalty to parties is reflected in increased ticket splitting (voting for Democrats for some offices and Republicans for others). It also may be reflected in a greater willingness to vote for a third-party or independent candidate, such as Ross Perot. Still, over the 1970s and 1980s the Democratic advantage in partisan loyalties eroded considerably, to the point that Republicans were close to parity among those who actually voted (Asher 1992, 86).

Summary

A number of attitudinal and social factors are related to individual voting behavior. Among attitudinal factors, assessments of the personal characteristics of the candidates, evaluations of the performance of the government, orientations on specific policy issues, and party identification are the primary determinants of candidate choice. For social factors, race, religion, region, social class, and gender appear to be the characteristics most closely related to voting. Examining how these factors are related to the vote in particular elections not only allows us to explain the election outcome, but also can provide us with an understanding of electoral dynamics. All of the ideas raised in this chapter can be examined, at least to a certain extent, with the data contained in this package.

Chapter III

Methods of Analysis

The study of voting behavior generally relies on information from sample surveys. Aggregate election statistics from states or counties, another common type of election data, are useful for examining the patterns of election results, such as differences in the presidential vote among the fifty states, but such data are not suitable for an analysis that focuses on the individual voter. In order to investigate the factors that affect how people vote, we need information on individuals. Such information commonly includes data on voting behavior, attitudes and beliefs, personal characteristics, and so on. Since it is impractical to obtain this information for each member of the electorate, the common procedure is to draw a sample of people from the population and interview these individuals. Once collected, survey data are usually processed and stored in a form allowing for computer-assisted data analysis. This data analysis generally focuses on describing and explaining patterns of political opinion and electoral behavior.

This chapter contains discussions of: (1) the nature of the dataset for the instructional package; (2) the data collection procedures for sample surveys; and (3) the basic principles for analyzing survey data. Although this is only a brief overview of these topics, this chapter should provide the basic information necessary to understand the analysis exercises contained in Chapter IV.

The Module Dataset

The data for this instructional package are drawn from the 1972 through 1992 American National Election Studies, conducted by the Center for Political Research at The University of Michigan. Based on very large samples (each on the order of 1,500 to 2,000 people), the studies interviewed respondents both before and after each presidential election. Since this dataset is a compilation of data drawn from six separate surveys, social scientists call the dataset a pooled cross-sectional dataset. It is cross-sectional since each election survey (1972 through 1992) can be thought of as a cross-section of the American electorate, meaning that the dataset is a representative sample of the American electorate at one given point in time. These cross-sections are then pooled—compiled into one large dataset.

Only a portion of all the information collected by the original studies is contained in the dataset supplied with this module. The selected data have been prepared especially for instructional purposes. The dataset, as a whole, contains data on over 13,000 individuals, collected in presidential years from 1972 through 1992. The American National Election Study typically asks several hundred questions of each individual surveyed. For this dataset, that large amount of information has been condensed into 70 variables that are of particular interest to data analysts.¹ In creating some variables, questions have been combined and data have been recoded, as will be explained below.

Efficient data analysis requires that the data be recorded, coded, processed, and stored according to standard procedures. Essentially, this involves representing all information by numeric codes. For example, the information that John Smith is a Catholic would be stored by recording a value of “2” (Catholic) on variable “67” (Religion) for respondent “907” (John Smith). This numerically coded information is placed on a magnetic computer diskette or tape, allowing the data to be analyzed with the aid of a computer. In the past, large surveys were analyzed with a “mainframe” computer, since the data analysis required a large amount of computer memory. Large survey datasets were thus often stored and distributed on magnetic tapes. The development of more powerful microcomputers now makes it possible for data analysis to take place on personal computers, in which case the data would normally be stored on a computer diskette or on a hard disk.

These instructional survey data can be thought of as a set of information available for each respondent included in one of the 1972 through 1992 surveys. This information indicates the placement of each respondent on a number of behavioral, attitudinal, and socio-demographic factors, as well as the year in which the data were collected. For example, this data set includes information on the respondent’s voting behavior (behavioral factors), the respondent’s attitudes about issues and candidates in the election campaigns (attitudinal factors), and the respondent’s gender, race, and education (socio-demographic factors). Each of these factors can be termed a variable. A variable is simply some factor for which there is a set of values (at least two) such that each person has some specific value (i.e., position or categorization) on that variable. Thus the respondent’s gender would be a variable and “male” and “female” would be the two possible values for that variable.

Some variables, such as the respondent’s gender or religious affiliation, clearly have a limited number of possible values. Other variables, such as the respondent’s age (expressed in years) can assume a great number of values. Of course, a variable that potentially has a large number of values can be made into a variable with a smaller number of possible values by recategorizing the responses. For example, this could be done to age by recoding respondents’ age into six categories as we have done in this module dataset. All of the variables in the dataset for this module have a small number of possible values. This was done to make the data easier to analyze. We will have more to say about recoding variables later.

One variable in this dataset is of particular interest—the year in which the data were collected. This variable is of interest since it allows the data analyst to look individually at any one election (e.g., 1972 or 1984) in which he or she might be interested; at any combination of elections that

12 Methods of Analysis

might prove interesting (e.g., 1980 through 1988); or the entire 1972 through 1992 period. The use of this variable will be illustrated below and will be thoroughly discussed in Chapter IV.

In order to use the dataset, a codebook is needed. The codebook describes the dataset by providing a list of all variables, an explanation of each variable, and a description of the possible values for each variable. The codebook also indicates how the data are stored and organized for use by the computer, including how each variable is coded. A codebook can thus be thought of as a combination of a map and an index to the dataset. The codebook for this dataset is located in Chapter VI. Detailed information for the use of the codebook is in Chapter V. This information should be read carefully.

Survey Research Methods

Analyzing the data for this module involves making generalizations and drawing conclusions about the American electorate from the set of people who were interviewed in the survey. Fortunately for social scientists, it is possible to study large populations by examining samples drawn from these populations. Of course, it is not possible to estimate perfectly the characteristics of a large population (e.g., American adults) even from a sample of thousands of individuals. Such a sample only provides a reasonably accurate representation of the large population. Because it is also possible for the data from a sample survey to be inaccurate, it is desirable to have some understanding of the potential sources of error in the data before engaging in data analysis.

The degree to which a sample is likely to represent some larger population depends on the sampling procedure. If we attempted to study all American adults by having interviewers stand on selected street corners and interview some of those passing by, then we probably would not obtain a sample that faithfully represents the entire population. First of all, the people who walk by street corners where the interviewers are may differ significantly from those who do not. Second, the interviewers may consciously or unconsciously choose certain types of people to interview. To improve our chances of obtaining a more representative sample, a different sampling plan is necessary, one that does not automatically exclude certain members of the population and does not allow for any selection bias from the interviewers.

The most basic form of sampling that conforms to the above principles is the simple random sample, which involves randomly drawing a sample from a list of all members of the population. For example, if we want a sample of 500 students from the undergraduate body of 15,000 at a large university, we could obtain the names of all undergraduate students from the university registrar, put each name on a slip of paper, put all the slips of paper in a hat (a *very* large hat!), shake the hat a few times, and draw a sample of 500 names from the hat. This procedure would guarantee that every individual in the population of 15,000 had the same chance of being included in the sample as every other individual, and that the choice of any one in-

dividual did not affect the chance that any other would be chosen. These are the keys to random sampling.

While simple random sampling is appropriate for some situations, it is not generally appropriate for most social science research applications. In the case of the university sample, we would probably want to make sure that the sample of 500 students contained relatively the same percentage of freshmen, sophomores, juniors, and seniors as were in the university undergraduate population of 15,000. A simple random sampling procedure would not guarantee this. By luck we might obtain a sample that had too many freshmen, or too few seniors.

A more complicated sampling procedure that *would* guarantee that the classes were represented according to their weight in the population is called stratified probability sampling. To take a stratified probability sample of the university, we would first divide the names obtained from the registrar into freshmen, sophomores, juniors, and seniors, and put each set of names into its own hat. We would then need to determine what percentage of the population fell into each of the four classes. Suppose we discover that 30% of the undergraduate population were freshmen, 27% were sophomores, 23% were juniors, and 20% were seniors (some college students do not make it all the way through to graduation while a smaller number transfer in from other institutions). To reflect accurately these percentages in the sample, we would need 150 freshmen, 135 sophomores, 115 juniors, and 100 seniors. The actual selection of the individual names for the sample would be conducted in the same manner as for a simply random sample. We would randomly select 150 names from the freshman hat, 135 from the sophomore hat, 115 from the junior hat, and 100 from the senior hat.

Modern survey research essentially applies these and other principles in its sampling procedures. Simple random samples are impractical in national studies for two basic reasons: (1) there is no national list of all American adults; and (2) the sample would be scattered all over the country, making it very expensive to send interviewers out to each respondent. For these reasons, a more complicated type of probability sample is normally drawn in national studies in such a way that the sampling procedure is as unbiased as the simple random sample.

The sampling procedure used to conduct a representative national survey depends on whether one is conducting a phone survey or a face-to-face survey. One commonly use method for drawing national samples for telephone surveys is to identify all of the area codes in use in the U.S., then identify all of the exchanges in use in each area code. After this is done, a computer is programmed to dial a four digit random sequence of numbers which is added to each of the telephone exchanges in each of the area codes. The actual number of respondents in any area code is determined by the actual number of telephone numbers assigned in the geographic area for which that area code is used. One major advantage of this “random digit dialing” technique is that it allows for people whose telephone numbers are unlisted in any telephone directory to be possibly included in a survey. A major disadvantage to pure random digit dialing is

the large number of calls that must be made to get a valid interview. Many calls are made to phone numbers that are unassigned; some are made to businesses. Typically, five calls must be made to get one working residential phone. Because of the large number of calls that need to be made to obtain valid interviews, most survey organizations modify pure random digit dialing and use sampling techniques that attempt to minimize this problem and so keep their costs within reasonable bounds. One widely used variation on pure random digit dialing that dramatically increases the odds of getting a residential phone number is known as the Waksberg method. The Survey Research Center at the University of Michigan uses the Waksberg method for telephone sampling.²

Compared to face-to-face interviews, telephone interviews have several disadvantages. (1) It is often difficult to get respondents to agree to a long interview over the telephone. This often means that interviews conducted over the telephone must be kept short. Thus fewer questions may be asked, follow-up questions dropped, or open-ended questions replaced with fixed answer choice items. (2) It is difficult to convey some kinds of questions over the telephone. For example, one commonly used device in the ANES is the feeling thermometer. In a face-to-face interview, respondents are handed a card with a picture of a thermometer reading from a low temperature of zero to a high temperature of 100. They are asked to rate their own feelings of warmth or coolness towards various political candidates, groups, or political parties. It is much more difficult to convey the concept of a feeling thermometer, and some other survey devices, over the telephone than it is in person. (3) Effective interviewing often involves the interviewer developing a rapport with the respondent so that the respondent feels comfortable in answering personal or intrusive questions. This is easier to do in a face-to-face situation than it is on the telephone.³

The chief advantages of telephone interviewing over face-to-face interviewing are cost and convenience. It is simply very costly to send interviewers out into the field to conduct interviews. Interviewers operating in the field on their own must be superbly trained to handle all kinds of interview situations. Telephone interviewing only requires a bank of telephones with interviewers at each one. All of the interviewers are in one place and can be carefully supervised by a well-trained supervisor. While face-to-face surveys are more difficult to conduct and more expensive than phone surveys, they usually yield richer data since the interviewer can ask a variety of follow-up questions and can note any circumstances about the interview that the interviewer thinks might have affected the answers given.

National samples for face-to-face surveys are drawn by an area probability method, which relies on U.S. Census figures and maps of the country. The American National Election Study uses a four-stage process to draw its national sample. This process, called a multi-stage probability sample, is based on a methodology developed at the Survey Research Center at The University of Michigan in the late 1940s. The first stage of this process for these studies involves selection of primary sampling units (psu's) from the

listing of all Metropolitan Statistical Areas (MSAs), single non-MSA counties, and groups of small non-MSA counties. Within each of these psu's, sample places are chosen and the number of interviews needed in each sample place is calculated. Typically, the largest population center in each psu would be a sample place. Then, either census blocks (in urban areas) or enumeration districts (in more rural areas) are selected randomly from each sample place. In the next stage, listings of all housing units located in the physical boundaries of the block or enumeration district are made, and a sample of housing units within each sampling unit is taken. In the last stage, all adults in a household are identified and one is selected on a random basis.⁴ The procedures for identifying who should be interviewed are objective; the interviewers are not allowed to replace somebody they are supposed to interview with somebody else. Because of the care taken in taking the sample and training interviewers, the only source of bias in this sampling plan would come from inaccuracies in the information supplied from the U.S. Census.

Regardless of how well designed the sampling procedure is, it is necessary to have a reasonably large sample in order to have confidence that the sample is representative of the population. Even if the finest sampling plan is used, a sample of ten people almost surely will be unrepresentative, as a result of random chance alone. But if a sample of 1,000 individuals is chosen by the same procedure, then it is very likely that the sample will be quite representative, as the random factors are likely to even out in a larger sample. Thus, we want to have both a well-designed probability sampling plan and a large sample size. Of course, a large sample will be more expensive than a small one, so interviewing costs are a major part of a survey.

The pooled cross-sectional nature of this module's dataset means that it is actually a combination of six different national samples, each identified in the year in which the survey was conducted. For all but the 1976 and 1992 surveys, the sampling procedures exactly followed those detailed above. In 1976 and 1992, two different sampling procedures were actually combined to select the sample for American National Election Study. Both 1976 and 1992 were years in which the ANES was part of a panel study. Panel studies are those in which the same respondents are interviewed several times over the course of the study in order to determine how their beliefs, attitudes, and behaviors have changed over time. For these two years, a portion of the national sample was cross-sectional and a portion was part of the ongoing panel. While the differences between the two samples identified in 1976 and 1992 should not overly concern you, some aspects of this procedure will be addressed below.

Interviews are never obtained from every individual selected in the sample. Some will refuse to be questioned, and more likely, others will not be at home when the interviewer calls. Interviewers can call back to obtain interviews from those who were not at home originally, but some will never be home when an attempt is made to contact them. The problem this represents is that those who fail to be interviewed may differ in some fundamental ways

14 Methods of Analysis

from those who are interviewed. A trivial example of this is a survey in which all interviews are planned for daylight hours. This would systematically exclude most working people. We might then expect to get a very different distribution of answers on questions about the importance of government programs to handle the problem of unemployment than if we had a survey in which some people were called in the evening. This potential source of bias will be a more serious problem when the non-response rate is high, so survey researchers try to minimize the number of non-responses. This usually means having the interviewers make repeated callbacks and interviewing at different times of the day, both of which increase the cost of the survey.

In panel studies, the potential bias as a result of non-response is a larger problem. First, non-response may be the result of a respondent moving during the course of the study. Interviewers may make heroic attempts to locate respondents who move, but it may simply be impossible to locate these people for further interviews. One relevant question thus becomes: Are people who move during a panel study different from those who do not move? While this is not a question that can be answered definitively here, it should be noted that there are some differences we might expect between movers and non-movers. Primary among these differences is that many more younger and single people might be expected to move over this short time period than would older and married people. A second reason that non-response might be a larger problem in a panel study is the nature of the study itself. Panel studies require respondents to be reinterviewed, often many times, something that a portion of the panel will refuse to do. A relevant question here becomes: Are those who are reinterviewed throughout the length of the panel different from those who refuse to be reinterviewed at some point? Research on this question has shown that people who last the length of the panel tend to be somewhat more interested in the subject of the study than those who refuse to be reinterviewed.

The process of obtaining information from the survey respondents by having interviewers administer questionnaires is another potential source of error. The respondents may be less than candid on some points, the interviewers may make mistakes, and some questions may be misleading. Thus, the information recorded by the interviewers may not be a perfect reflection of the true positions of the respondents. However, if a well-designed questionnaire and well-trained interviewers are used, this type of error can be held to a minimum. Skilled interviewers are less likely to make careless errors in asking questions and recording information. A well-designed questionnaire will have each question worded in a clear and unbiased manner, which will help to produce accurate results. Again, minimizing error involves some cost. Extensive prior testing of the questionnaire before actual use involves expenditures of time and money, as does careful selection and training of interviewers.

It is clear that while the principles for obtaining a good survey are well known, practical considerations place limits on the application of these principles. A survey research organization generally has a fixed sum of money for conducting a particular study, and this financial limitation may

dictate that some procedures be done in less-than-ideal fashion. High quality surveys will cost more, but the information collected will be more accurate.

The data for this instructional package are drawn from a series of academic studies that are considered to be of high quality in all aspects, including the sampling plan, response rate, questionnaire design, and so on. All the respondents were interviewed in the fall of the year before each presidential election. Almost all of these were interviewed in person; a small proportion were interviewed by telephone. All interviews were conducted by highly trained interviewers. Reinterviews were attempted with each respondent after the election, although some could not be reached. This dataset includes respondents interviewed in both the pre-election phase (September or October) and the post-election phase (November or December) of each year.

In the 1976 and 1992 studies, it was found that it was somewhat easier to recontact those who were originally in the panel portion of the sample than it was to recontact those who were originally in the cross-sectional portion, thus resulting in a potentially biased sample when the two portions were combined. A correction for this situation was made to the dataset in those two years. This correction is called a weight. Weighting a dataset is a technical procedure in which each respondent counts as more or less than one, depending on the value of the weight assigned to that respondent. If members of the 1976 or the 1992 panel sample had a higher probability of being reinterviewed than the remainder of the sample, then the panel members have a greater chance of being included in the responding sample. This was corrected by decreasing the weight of each respondent in the panel and increasing the weight of each respondent who was not in the panel. The 1992 data are also weighted to correct for other factors, such as the fact that individuals in different size households had different probabilities of being selected for interviews.⁵ The resulting weighted data set contains 13,974 rather than 13,349 respondents.⁶

Even in surveys of high quality, some error is present. Some of the error will essentially be random in nature, resulting primarily from sampling error (the error introduced by the sampling procedure itself) but also from accidental error by interviewers or coders. In addition to the random error, there will be systematic error. Certain questions may have an unrecognized bias to them, so that respondents tend to respond in a certain way even if they feel differently. Another possibility is some systematic bias to the sample, for any sampling plan tends to exclude certain remote and inaccessible types of people. For example, the Survey Research Center sampling plan used by the American National Election Study excludes residents of Alaska and Hawaii and people living on military bases.⁷

The presence of random error means that the results obtained by the study must be taken as close approximations of the population rather than perfectly precise representations. For example, if 60% of the people interviewed in the study said that they were in favor of spending more money on defense, it would be incorrect to conclude that exactly 60% of all American adults felt that way. The true popu-

lation figure should be about, but probably not exactly 60%. For a sample of about 2200 respondents, the random error introduced by the process of sampling alone will be less than 2 percentage points 95% of the time; for a sample of about 1900 respondents, the random error increases to 3%, and for a sample of about 1500, the random error increases again to 4%. In other words, if our sample shows that 60% of the people surveyed in 1992 favor increasing expenditures on defense, then we would be very confident that the true population figure is between 58% and 62% (i.e., 60% minus 2% and 60% plus 2%). For the same question asked in 1980, when the sample was smaller than in 1992, we would be confident that the true population figure is between 57% and 63%. Sampling error is dependent on sample size; it is larger for a smaller sample, or when a subset of a larger sample is used, and less for a larger sample. The existence of random error means that we must pay attention to the total number of respondents that our conclusions are based on.

From a sampling purist's point of view, the data in this module dataset are not one sample but a compilation of six samples. For that reason, it is difficult to identify what the sampling error would be for the dataset taken as a whole. If we could consider the 13,000+ cases in this dataset as *one* sample taken from the U.S. population as a whole, that sample would certainly have a very small sampling error. But that is *not* the case with these data; although the sample methodology is similar from year to year, the samples *were* taken over a number of years with the possibility that some members of the population *could* have been drawn for more than one study. This should not be a major problem as it is not very often that the dataset will be analyzed in its entirety—most often we will be seeking to compare the results obtained for an analysis of one election with those obtained for another election (e.g., comparing the relationship between party identification and presidential vote in 1976 and 1984). When the dataset is analyzed as a whole, the results we obtain should be thought of an estimate for the six elections under study, that is an average figure for the six elections.

However, even this discussion of random error begs the point, for we are less interested in the proportion of respondents who favor a given government program than we are in the *relationship* between one question and others. This will be discussed in greater detail below; suffice it to say here that social science researchers, unlike pollsters, are more interested in understanding why people voted the way they did than they are in predicting the exact proportion of the population who will (or did) vote for a given candidate for office.

Systematic error presents different problems. While many of the questions generally asked in surveys may not produce data that contain significant systematic error, some questions have the potential to do so. For example, if a survey of all American adults finds that 5% of the respondents stated that they cheated on their income tax, it would be foolish to conclude that in reality only one American in twenty engages in this practice, for there is the very real possibility that many of the respondents did not answer the

question truthfully. Other types of systematic error occur when respondents are not consciously being untruthful but simply have forgotten what a past behavior was, or when respondents mistakenly interpret a survey as some kind of test of their knowledge of or involvement in politics. For example, voting studies in the 1950s routinely asked respondents to recall whom they had voted for in presidential elections ten or twenty years earlier. These questions were discarded when it was found that a large number of respondents could not remember their true past behaviors. Another example also comes from early voting studies which found that a substantial number of respondents changed their opinions on public policy issues over a short period of time; further investigation showed that many of these respondents really did not have opinions but gave answers anyway since they did not want to appear uninformed to the interviewer (see Converse 1964; 1970; Converse and Markus 1979). A last type of systematic error is the result of poor question wording. Poorly worded questions can encourage respondents to express an opinion on an issue when they do not have one. The best way to minimize this type of error is for experts in the fields of public opinion and survey research to review questions for content, presentation, and interpretation. The American National Election Study uses this method of survey design. Nevertheless, there really is no way to know exactly how much systematic error, if any, is present for a specific item. We should, therefore, be sensitive to the fact that on some types of questions respondents will not be truthful or accurate. The problem is especially important when one answer to a question appears to be more socially undesirable.

In sum, it is important to be aware of the potential sources of error when working with survey data. Fortunately the error present in a well designed and conducted survey will be minimal in most cases. It therefore is possible to draw meaningful and valid conclusions from an analysis of survey data. At the same time, it is necessary to realize that there can be some differences between the true behavior and opinion of a population and what the survey shows the behavior and opinions to be.

Principles of Data Analysis

Reports of surveys or polls in the popular media describe behavior or opinion, usually reporting such things as the proportion of people who favor some proposal, who feel a certain way, or who have engaged in some activity. These are often called the “what” questions of social science research—What does the American public think about abortion? Research in political science is concerned with more than mere description of data. The emphasis is on analysis and explanation, or “why” questions—Why did the Republican party win four of the six presidential elections from 1972 to 1992 when there have been more people who call themselves Democrats than there are Republicans over that same time period? Such questions involve examining relationships between variables in addition to the distribution of answers on the variables.

Two variables are related to each other when certain values of one variable are likely to be associated with certain

16 Methods of Analysis

values for the other variable. For example, if we say that education and turnout (i.e., whether one votes) are related, this could mean that more educated people are more likely to vote (which is called a direct or positive relationship), or that more educated people are less likely to vote (which is called an inverse or negative relationship). Either way, the two variables would be related, for differences on one variable would be linked to differences on the other variable. Naturally, it would be far more informative to state how education and turnout are related, whether positive or negative, rather than just state the simple fact that they are related, and this should be done whenever possible. Conversely, two variables are unrelated when a value of one variable is equally likely to be associated with any of the values for the other variable.

When speaking about the relationship between two variables, the terms independent variable and dependent variable are commonly used. The independent variable can be considered to be the “cause” and the dependent variable the “effect.” In other words, the independent variable affects or influences the dependent variable. A common research procedure is to start with some dependent variable and then to identify some independent variables that are strongly related to the dependent variable. In this way the dependent variable is explained, at least in the sense that some of the factors that influence it are identified.

When analyzing survey data, a common procedure is to use a contingency table. A contingency table presents the cross-tabulation between two variables. (Tables with three or more variables are possible, but we shall refer only to the more basic table for now). A cross-tabulation refers to the pattern of joint occurrences for the two variables. By examining a contingency table, it is possible to determine whether or not two variables are related.

For example, Table III-1 is a contingency table that presents the cross-tabulation between party identification and presidential vote for all six years combined. One variable, the respondent’s party identification (divided into seven categories) is at the top of the table, with one column for each of the seven values for the variable. The other variable, presidential vote, is along the side of the table with one row for each of the three values of the variable (those voting for minor party candidates are excluded from this table). The seven columns and three rows intersect to form twenty-one cells.

The twenty-one cells of Table III-1 correspond to the possible categories that a respondent could fall into. A respondent could be a strong Democrat who voted for the Democratic party’s candidate, or an independent Democrat who voted for the Republican party’s candidate, and so on. The number in parentheses in each cell indicates the total number of respondents who fall into the category (e.g., there were 1119 respondents who were strong Republicans and who voted for the Republican party’s candidate for president). The numbers on the right side and the bottom of the table are called marginals, and they indicate the total number of respondents in each row or column. The number in the lower right corner ($N = 8463$) is the total number of respondents included in the table.

Respondents who did not vote and/or did not respond to the party identification question are not included in this table, so the total is somewhat less than the 13,972 in the sample. These 5510 individuals all show missing data on one or both of the variables that compose Table III-1. Missing data occurs in the interview situation because the question does not apply to the respondent (e.g., people who did not vote were not asked which presidential candidate they voted for), the respondent refused to give an answer or had no opinion on the issue, or the interviewers failed to obtain the information for some other reason. Missing data can also be the result of a recoding scheme in which people who do not fit the categories that the researcher is interested in are recoded to missing. Table III-2 shows one example of this where third party voters are recoded as missing. This table thus gives us the relationship between party identification and the two-party vote. The 1980 and 1992 elections were interesting ones because of the presence of strong third candidates—John Anderson in 1980 and Ross Perot in 1992. In such elections, researchers are often interested in looking both at the vote for all three candidates and at the two-party vote (the vote for only the Democratic and Republican party candidates).

In order to interpret contingency tables, it is desirable to convert the cell frequencies into percentages. The usual way to calculate percentages in a contingency table is by each category of the independent variable. The independent variable is typically placed at the top of the table. Since party identification clearly is the independent variable in this case (we are interested in how party identification affected the vote; not how the vote affected party identification), the percentages have been calculated by column (i.e., each column totals to 100%). For example, Table III-1 shows us that 69.6% of the 960 independent Democrats voted for the Democratic party candidate, while 20.5% voted for the Republican party candidate, and 9.9% voted for third party candidates. According to the data in Table III-1, strong Democrats are the most likely to vote for Democratic party candidates and strong Republicans the least likely, with the other groups basically arranged between these two extremes.⁸ While the overall tendency is clear, we should not treat each of the percentages as a perfect reflection of the true population figure, since we are dealing with a sample. Thus, it would be better to say that, for the country as a whole, *around* 70% of the independent Democrats cast ballots for the Democratic party candidate for president during the past six years.⁹

This simple example illustrates the basic ideas behind reading and interpreting contingency tables. The general rule is to compare categories of the independent variable in terms of the percentage distribution on the dependent variable. If the independent variable is the column variable (as in Table III-1), then percentages should be calculated by column and the entries in the columns compared.¹⁰ A proper observation drawn from Table III-1 is: 87.6% of the strong Democrats voted for Democratic party candidates for president while 94.3% of the strong Republicans voted for Republican party candidates in presidential elections from 1972 to 1992. This indicates that strong Republicans were

TABLE III-1: PRESIDENTIAL VOTE BY PARTY IDENTIFICATION

<i>Pres. Vote</i>	<i>Strong Dem.</i>	<i>Weak Dem.</i>	<i>Indep. Dem.</i>	<i>Indep.</i>	<i>Indep. Rep.</i>	<i>Weak Rep.</i>	<i>Strong Rep.</i>	
Democrat	87.6% (1374)	63.9% (1093)	69.6% (668)	33.9% (249)	11.7% (118)	12.3% (159)	3.1% (37)	43.7% (3697)
Republican	10.9% (165)	31.5% (539)	20.5% (197)	55.2% (406)	80.8% (816)	81.2% (1051)	94.3% (1119)	50.7% (4292)
Other	1.9% (30)	4.6% (78)	9.9% (95)	10.9% (80)	7.5% (76)	6.5% (84)	2.6% (31)	5.6% (474)
	100.0% (1570)	100.0% (1710)	100.0% (960)	100.0% (734)	100.0% (1010)	100.0% (1294)	100.0% (1186)	N = 8463

more “loyal” to their candidates than strong Democrats were to theirs.

In comparing column entries, the focus should be on looking for differences, and one should note the size of the difference identified. Very small differences should be treated as insignificant, on the grounds that these small differences could be the result of random error and are probably not meaningful anyway. Where the differences are large enough to be considered real, it is important to note how strongly the variables are related. For example, in Table III-1 we can see that the differences among Democrats, Republicans, and independents in their presidential voting is substantial. The difference in support for Democratic presidential candidates between the column labelled “Strong Democrat” and the column labelled “Strong Republican” is nearly 85 percentage points (87.6% – 3.1% = 84.5). This is one of the largest differences in percentages that you will see in this dataset. If the difference between strong Democrats and strong Republicans was only about ten percentage points, we would classify the relationship as a weak one. And if there were virtually no differences at all in the percentage distributions in each column, then we would say that no relationship existed.

We also should pay attention to the total number of respondents that the column percentages are based on. If there are too few cases in a column, the percentages in the column cannot be regarded as reliable. As a general rule, we should be cautious in interpreting percentages when there are fewer than 50 respondents in the column (note that we are referring to the column marginal, not the cell!), and when there are fewer than 20, we should be extremely cautious.

To measure the strength of relationships more precisely, statisticians have developed various measures of association that can be calculated for a table. Many different measures of association have been designed, but they all have some general characteristics: they are equal to zero when there is no relationship, and they increase as the relationship becomes stronger, until they reach a value of 1.00 for a perfect positive relationship (or –1.00 in some cases where a perfect inverse relationship exists). However, you will not encounter any perfect relationships with these data. The relationship in Table III-2 is one of the strongest to be found in the 1992 data.

One commonly used statistic is Kendall’s tau. This measure of association is appropriate when we have variables that are ordinal, meaning that their categories form a logical order or progression, such as from high to low or from liberal to conservative. (Any variable with only two categories can be considered ordinal.) For example, if we had a table showing the relationship between education (an ordinal variable) and turnout (which has two categories), tau would be an appropriate statistic to determine the strength of the relationship between the two variables. It would equal zero if more educated and less educated individuals voted at the same rate, and it would equal one if all the more educated people voted and all the less educated people did not vote. For situations between these two extremes, tau would be somewhere between zero and one. For larger tables (tables with, for example, four columns and four rows), tau will equal +1.0 only when all cases are on the major diagonal of the table—that is, the only cases in the table are in the cells that start in the upper left hand corner of the table and proceed diagonally to the lower right hand corner. Rectangular tables (where the two variables have different numbers of categories) do not have clear diagonals. For this reason, some statistical packages report Kendall’s tau-*b* to be used with square tables and Kendall’s tau-*c* to be used with rectangular tables. Other statistical packages report only tau-*c*, since the result of calculating tau-*c* for a square table will be very similar to tau-*b*.

When we have ordinal variables, we frequently refer to positive and negative relationships. A positive relationship exists when the changes in the independent variable are associated with same direction changes in the dependent variable. For example, if better educated people display a higher turnout rate, the relationship would be positive. A negative relationship exists when we have the reverse; changes in one variable are associated with opposite direction changes in the other variable. Tau and other statistics appropriate for ordinal variables will assume negative values in these cases.

Table III-2 illustrates these points. The table displays the two-party presidential vote by party identification. Party identification is an ordinal variable and the two-party presidential vote is a variable with only two categories. Since two category variables can be assumed to be measured on an ordinal-level scale, tau-*c* is thus an appropriate statistic

TABLE III-2: TWO-PARTY PRESIDENTIAL VOTE BY PARTY IDENTIFICATION

<i>Pres. Vote</i>	<i>Strong Dem.</i>	<i>Weak Dem.</i>	<i>Indep. Dem.</i>	<i>Indep.</i>	<i>Indep. Rep.</i>	<i>Weak Rep.</i>	<i>Strong Rep.</i>	
Democrat	89.3% (1374)	67.0% (1093)	77.2% (668)	38.0% (249)	12.6% (118)	13.2% (159)	3.2% (37)	46.3% (3697)
Republican	10.7% (165)	33.0% (539)	22.8% (197)	62.0% (406)	87.4% (816)	86.8% (1051)	96.8% (1119)	53.7% (4292)
	100.0% (1539)	100.0% (1632)	100.0% (865)	100.0% (654)	100.0% (933)	100.0% (1210)	100.0% (1155)	N = 7989

$\tau_{uc} = .72$

to be calculated for this table. The highly positive tau-*c* calculated for Table III-2 can be interpreted as meaning that there is a strong relationship between party identification and the two-party presidential vote—as one identifies more strongly with the Republican party, one’s tendency to vote for Republican candidates increases.

When the variables are nominal (i.e., there is no logical order to the categories), either Cramer’s *V* or the contingency coefficient are appropriate measures of association. For example, if we had a table showing the relationship between religious affiliation and presidential vote, Cramer’s *V* would be appropriate; it would equal one if we could predict the presidential vote perfectly simply by knowing the person’s religion (e.g., if all Protestants voted one way, all Catholics another way, and so on) and it would equal zero if each religious group voted in the same manner. The contingency coefficient could also be calculated for this table and would yield similar results.

Look back at Table III-1. Since the three-candidate presidential vote can be considered to be a nominal level variable, Cramer’s *V* and the contingency coefficient are appropriate statistics to summarize the strength of the relationship in the table. For this table, Cramer’s *V* = .47 and the contingency coefficient = .56. These are high levels for either of these statistics to display, indicating that there is, indeed, a strong relationship between party identification and the presidential vote.

In addition to tau, Cramer’s *V* and the contingency coefficient, there are many other statistics for use with contingency tables. Commonly used statistics for ordinal variables are gamma and Somer’s *d*, while lambda is another commonly used statistic for nominal level variables. We shall leave it up to your professor to suggest which measures of association, if any, you should use. The great value of these statistics is that they can summarize in a single number an entire table with many cells and percentages. On the other hand, a disadvantage of summary statistics is that they summarize the relationship. In doing so, some detailed information about the relationship is lost.

Remember that measures of association do not really tell you any more than percentages in the table, they just do it more conveniently. Also remember that most measures of association can be calculated on any data you might have. It makes little theoretical sense to measure the relationship between the incidence of sunspots in any given election

year and the results of the presidential election in that year, and yet there is nothing in any computer routine that would stop you from doing so! Nor do computer routines stop you from calculating ordinal level statistics on data that are nominal, or vice versa. Whether the measure you select is appropriate (and therefore useful) for the data you calculate it on, and whether the variables you choose to measure relationships between make theoretical sense, are matters of your own substantive knowledge of voting behavior and methodological training in data analysis.

Additionally, you should bear in mind that the sign for a measure of association you ask the computer to calculate for you depends on how the variables in your table are coded. If, for example, you cross-tabulate education with turnout and ask the computer to calculate tau, tau will be positive if the variables are coded low to high. In other words, if education were coded 1 = Not a High School Graduate; 2 = High School Graduate; 3 = Some College; 4 = College Graduate and turnout were coded 1 = Did Not Vote and 2 = Voted, then tau will be positive. If, however, one *or* the other of the two variables is coded high to low (1 = College Graduate; 2 = Some College; etc.), tau will be negative. Remember, a positive tau means that as one variable increases, the other also increases; a negative tau means that as one variable increases, the other decreases.

Computer-based Statistical Packages and Data Analysis

The survey data for this instructional package can be analyzed according to the above principles. The specific computer commands and methods for generating the desired contingency tables will depend on the local computing environment, but most students will be using one of two basic methods provided by this instructional package. One option is to rely on a mainframe computer and use SPSS, a popular computer-based statistical package. The other choice is to use microcomputers and use SPSS for Windows or for the Apple MacIntosh, popular interactive microcomputer statistical packages.¹¹

In SPSS or SPSS for Windows (both of which we will abbreviate as SPSS from now on), only a small number of commands will be necessary to generate the desired analysis. Some of these commands will be specific to your particular computer system and will have to be provided to

you locally. Other commands are independent of the particular system and can be explained here. The *crosstabs* command would be used to generate the tables in the exercises in Chapters III and IV of this monograph. When using *crosstabs*, you should remember that the variables are specified as row variable by column variable. In this monograph, we adopt the convention of making the row variable the dependent variable and the column variable the independent variable. When a table is formatted in this way, we should specify that column percentages be calculated for the table.

The command used to generate Table III-1 is:

crosstabs tables=v04 by v09/cells=counts col.

This command shows the basic syntax of *crosstabs*. In addition to specifying the table through the use of the variable numbers (V04¹² is vote in presidential elections; V09 is party identification), the commands tells SPSS to print the actual number of cases in each cell of the table (counts) and to print the column percentages (col). If you are using SPSS for Windows, you can generate this table by typing this command in the syntax window (and clicking on the “run” icon) or you can generate this table by clicking on the proper menus.

Recoding Variables

Although the variables in this instructional dataset have been coded into what we think are the most logical categories, there are times that you might want to change the coding. For example, we have coded education (V62) to have four categories (Not a HS graduate; HS graduate; Some college, and College graduate). Although these represent logical categories, there are at least two good reasons why one might want to recode the education variable. First, you may think that the real distinction in education is between those people who have not gone to college and those who have (even if people who went to college did not graduate). Second, you might find that there were too few cases in one of the four categories to be able to make valid conclusions about some difference you find between columns in a table, especially if you were looking at only one presidential election (e.g., 1980). If either of these were the case, you could recode the education variable into two categories instead of four by “collapsing” those who have a “1” or a “2” on education into “1” and renaming this category something like “HS Only.” Those with a “3” or a “4” on education could then be collapsed into a “2” and renamed something like “Any College.”

Recoding can have theoretical overtones. Let us examine party identification (V09). V09 is coded here on a seven-point scale which runs from Strong Democrat to Strong Republican with independent being the middle category. For some uses, the seven-point scale has too many categories and results in tables that are too complex. In other cases using the seven point scale might result in tables with too few respondents for meaningful analysis in some categories. Therefore, recoding V09 into a smaller number of categories might be desirable here. There is some controversy, however, about exactly *how* party identification

should be recoded. By consulting the codebook you will see that party identification is assessed through answers to three questions: First respondents are asked whether they generally consider themselves to be Democrats, Republicans, or independents. For those who call themselves Democrats or Republicans, the strength of their attachment to their party is then asked. Those who call themselves independents are asked whether they feel closer to the Republicans, Democrats, or neither.

Three categories of party identification (Democrat, independent, Republican) would result in large numbers of respondents in each category, but at the sacrifice of knowing the strength of the respondents’ attachment to their political party. Nevertheless, we could accomplish such a recoding easily enough, as long as we decided beforehand who is a Democrat, who is an independent, and who is a Republican. The particular problem here concerns those who originally called themselves independents but then say that they are close to either the Democratic or Republican party. There are two alternative treatments that can be used here. First, we could stay with the initial question sorting and treat all those who originally called themselves independents as independents in the recoded variable. Our coding scheme would thus combine all those who are originally a “1” or “2” into “1” (Democrat), all those who are originally a “3,” “4,” or “5” into “2” (independent), and all those who are originally a “6” or “7” into “3” (Republican). Or second, we might remember the results we obtained in Table III-1 and note that independent Democrats behave very much like weak Democrats, and independent Republicans much like weak Republicans. We might then want to combine independent Democrats in with the weak and strong Democrats, and the independent Republicans in with the weak and strong Republicans. This second coding scheme combines all those who are originally a “1,” “2,” or “3” into “1” (Democrat), all those who are originally a “4” into a “2” (independent), and all those who are originally a “5,” “6,” or “7” into a “3” (Republican). Needless to say, there would be large changes in the numbers of respondents in each of the three categories depending on which coding scheme was used. You can see from this example that the jury is still out on how party identification should be recoded.¹³

Recoding is a fairly easy process in SPSS or most other computer-based statistical packages. In SPSS, we could recode party identification (V09) in the first way proposed above with the following commands:

```
compute v09r1=v09.
recode v09r1 (1,2=1) (3,4,5=2) (6,7=3).
variable labels v09r1 ‘Recoded Party ID Leaners Ind’.
value labels v09r1 1 ‘Dem’ 2 ‘Ind’ 3 ‘Rep’.
```

This series of commands does several things. First, it creates a new variable (V09r1) which is the equivalent of V09. V09r1 is then recoded into three categories. You should note that immediately following the *recode* command, we have assigned a variable label to V09r1 and have assigned value labels for party identification.

These commands show several good data analysis habits to develop. First, create a new variable to recode, rather than recoding the original variable. When an original variable is recoded, it remains recoded for the rest of the computer run in which you make the change. Thus, it is impossible to “recapture” the original unrecoded variable in the same computer run, if that variable has been recoded. Creating the new variable allows you to keep and work with the original unrecoded variable while also working with the new recoded one. Second, assign variable labels and value labels to new variables as soon as you create them. It is often clear what you want to do in a computer run when you are doing it, but later when you examine the printout from the run, your intentions are less clear. Naming a variable when it is created and assigning value labels immediately after creating the new variable will help you to keep things straight. If you do not add variable and value labels when you create new variables, you may not be able to tell what these variables are.

For the second party identification recoding scheme described above, the SPSS statements would look very similar to that above, but would be:

```
compute v09r2=v09.
recode v09r2 (1,2,3=1) (4=3) (5,6,7=3).
variable labels v09r2 'Recoded Party ID Leaners Part'.
value labels v09r2 1 'Demo' 2 'Ind' 3 'Rep'.
```

You should note here that, although the value labels are the same as those in the first recode command above, the number of cases that will go into each category will be different. Specifically, here we are calling more respondents Democrats and Republicans (and fewer respondents independents) than we do above. You should also note that we have assigned somewhat different variable labels to the two new variables (V09r1 and V09r2) as an extra aid in keeping track of the new recoded variables.

There are a variety of reasons beyond collapsing categories why one might want to recode variables. One major reason might be to change the order of the codes that were

assigned to a variable. In Table III-1 (above), we noted that the codes for Presidential Vote (V04) were 1 = Democratic Party candidate, 2 = Republican Party candidate, 3 = Other party candidate. One might develop a theory that voting for other party candidates is some sort of compromise for individuals who do not particularly like their own party candidate but who simply cannot bring themselves to vote for the other major party candidate. So a Republican who was disaffected with President Bush's handling of the economy might, in the 1992 presidential election, vote for Ross Perot, because he or she simply could not vote for the Democratic Party candidate, Bill Clinton. This theory would require that v04 be recoded 1 = Democratic Party candidate, 2 = Other party candidate, 3 = Republican Party candidate. This type of recoding can be easily accomplished with a series of statements like the following:

```
compute v04r=v04.
recode v04r (1=1) (2=3) (3=2).
variable labels v04r 'Recoded Presidential Vote'.
value labels v04r 1 'Dem' 2 'Other Party' 3 'Rep'.
```

The intention of this series of statements is to change the order of the categories in order to create an ordinal-level scale.

Conclusion

The necessary instructions for using computers at your school to analyze the data supplied with this instructional package will be supplied by your professor. Your school's computing environment and your professor's preferences about computer-based instructional packages will determine which package you use to work on the exercises in Chapter IV. Although we have used SPSS to illustrate the commands for cross-tabulation and recoding, this is certainly not the only computer-based statistical package that can be used to complete the exercises in Chapter IV. If SPSS is not available at your school, the commands for cross-tabulation and recoding will be different than those shown above. In this event, your professor will supply you with the proper commands to complete the exercises.

Notes

1. One of the major considerations in our choice of questions to be used in creating this dataset is the presence of the same question (or questions with very similar wording) in each of the ANES questionnaires from 1972 until 1992. Only in rare occurrences did we choose questions for the dataset which were not asked in each of these election years.
2. For a good description of telephone sampling and of the Waksberg method of sampling, see Weisberg, Krosnick, and Bowen (1989), pp. 43-49.
3. There are many recent studies that seek to compare the results of interviews conducted in person with those conducted on the telephone. For those interested in further reading on this subject, some suggestions are Rogers (1989) and Groves (1989).
4. See Weisberg, Krosnick, and Bowen (1989) for a good description of how the Survey Research Center draws its sample. The methodology for drawing this sample was originally developed in Leslie Kish, former director of the Survey Research Center. See Kish (1949).
5. An example might make weighting easier to understand. Many surveys use random digit dialing to choose a sample of respondents to be interviewed by telephone. A random list of telephone exchanges for a given city is identified and then a random four digit number is attached to this exchange. A computer dials the numbers one-by-one and when a person answers the phone, the interviewer determines if an interview should be conducted. It should be apparent that the sample that has been isolated in this example is *not* a sample of people but of telephone numbers. Many people have more than one telephone number; some have three or four. These people have twice or three times or four times the chance of being included in the sample as somebody who only has one telephone number. Researchers who conduct telephone interviews using random digit dialing have thus taken to asking respondents how many telephone numbers they have and weighting their responses by the reciprocal of the number of telephone numbers in the household (1/2 for two phones; 1/3 for three; etc.). This “corrects” the resulting dataset so that it resembles a good sample of individuals.
6. The increase of 625 respondents comes almost solely from weights assigned to the 1976 data. The unweighted number of cases for 1976 is 2248 while the weighted number is 2870. The unweighted number of cases in 1992 is 2485 while the weighted number is 2488.
7. There is one additional source of potential error in this pooled cross-sectional dataset that needs to be discussed in detail here—the process of pooling the data might introduce its own source of error. This can occur as a result of the different number of cases in each of the survey years. We should expect that the number of respondents in each of

the six survey years would be about 1/6 (16.7%) of the total number of respondents. For the unweighted data, the actual range of percentages runs from a low of 12.1% in 1980 to a high of 20.3% in 1972. When the data are weighted to make the 1976 and 1992 samples representative of the electorate, the actual range of percentages changes from a low of 11.6% in 1980 to a high of 20.5% in 1976. The large number of weighted cases in 1976 is a result of the weighting procedure used in that year.

The important point here is that the 1972 and 1976 data, added together, represent 39.6% of the total dataset, when they “should” only represent 33.4%. This potential source of error only becomes troublesome in a small number of instances—in particular, when one is interested in looking at the entire dataset as a whole *and* one suspects that the early data (1972 and 1976) somehow differ from the later data (1980 through 1992). However, the nature of the dataset makes it likely that separate analyses would be run for the 1972 and 1976 samples, on the one hand, and the 1980 through 1992 samples on the other.

One way to correct for the above is to re-weight each of the cases so that each year counts for 16.7% of the total. We have provided such a weight variable in the dataset and some instructors may choose to use this option. Others may choose not to do so because they wish to keep the number of cases in any one year equal to the original survey.

The tables and exercises we present in this book do *not* re-weight the data to make each year equal to 16.7% of the sample. If your instructor chooses to re-weight the data, all tables contained herein will need to be re-calculated.

8. You should note that independent leaning Democrats show a stronger tendency to support Democratic presidential candidates than do weak Democrats (69.6% compared to 63.9%) and independent leaning Republicans show a stronger tendency to support Republican presidential candidates than do weak Republicans (81.2% compared to 80.8%). This is a well-documented tendency in voting research. For possible explanations, see Wattenberg (1994).

9. Table III-1 shows one of the odd characteristics of working with weighted data. Examine the column labeled “Strong Democrat” in the table. This shows that 1374 Strong Democrats voted for Democratic party candidates, 165 voted for Republican party candidates, and 30 voted for other party candidates. The column marginal (the total number of Strong Democrats in the table) is 1570. But 1374+165+30 equals 1569 not 1570! This is because of rounding error. Usually, rounding error is discussed in conjunction with percentages—sometimes the percentages in a column add only to 99.9% or to 100.1%—because the computer program rounds to the nearest tenth of a percentage point. If there were nine respondents in a column of three rows, and each cell in the column had three respondents in

22 Methods of Analysis

it, the column percentages would be 33.3% for each cell. These add to 99.9% rather than 100.0%. With weighted data, the situation is analogous, but the rounding error concerns *respondents* rather than percentages. The computerized data analysis package used to generate the tables in this monograph (SPSS) rounds weighted respondents to the nearest whole person *but calculates percentages on the unrounded respondents*.

Examine again the Strong Democrat column in Table III-1. Some 87.6% of the Strong Democrats—or 1374 individuals—voted for Democratic candidates in the presidential elections from 1972 to 1992. But 87.6% of 1570 does not exactly equal 1374. The number of individuals in this cell should be approximately 1375.32; we say approximately because 87.6% is a rounded percentage which could be as low as 87.56% or as high as 87.74%. However, most of us are a bit uncomfortable dealing with .3 person, so SPSS rounds this off to a whole number, but calculates the percentage in the cell on the unrounded number. Likewise, the number of Strong Democrats voting for Republican candidates should be about 164.85 rather than 165, and the number of Strong Democrats voting for other parties' candidates should be about 29.83 rather than 30. If we add $1375.32 + 164.85 + 29.83$, we get 1570. So properly rounding the individuals in each cell results in something that looks strange—the column marginal is not the addition of

the cell entries in that column—but it really is not strange at all.

Other computerized statistical data analysis packages, such as SAS, handle weighted data differently than SPSS does and will produce a table that shows fractions of a respondent in any given cell, if the rounding procedure actually produces such fractions.

10. Of course, if the independent variable is the row variable, the percentages should be calculated by row and row percentages should be compared with each other. In this monograph, we will always format tables with the independent variable at the top, calculate column percentages, and make comparisons between columns.

11. The commands for all SPSS products are very similar but there are some technical differences between the packages. For this reason, the examples listed below are written using commands we know will work in various versions. There may be more efficient ways of accomplishing the same ends in various SPSS packages; we leave these up to your professor and you to develop.

12. Note that the variable is V04 not V4.

13. For a fuller discussion of the concept of party identification and the controversy surrounding its meaning, see Wattenberg (1994).

Chapter IV

Analysis Exercises

In the last chapter, the computer commands necessary to generate Tables III-1 and III-2 were described. In order to proceed with this instructional package, it will be necessary for you to generate your own tables in like manner as Tables III-1 and III-2 were generated. The first two exercises in this chapter will allow you to make sure that you can properly obtain the tables you want; you can check your results against the tables in this chapter to verify your understanding of the appropriate computer procedures. The exercises also will require you to apply the principles of table reading that were discussed in the previous chapter. These elementary exercises will then be followed by more complex ones that involve more sophisticated techniques. Last, we will suggest some further projects that you might want to conduct on your own.

The pooled cross-sectional nature of this dataset allows for the data to be used in a number of ways. First, we can look at the relationship between two variables in any one electoral year. Second, we can look at the relationship between any given variable and the electoral year in order to determine whether there have been changes across the years. And third, we can use the dataset for several years or as a whole in order to look at subsamples of the respondents that would typically be too small in number in any one year to make in-depth analysis possible. These three uses of the dataset are illustrated by the exercises below.

Exercise 1: Ideology and Partisan Choice

In the following exercise, you will be creating tables that are analogous to Tables III-1 and III-2. We will begin by looking at the relationship between presidential vote and the respondents' ideological orientation.

Exercise 1A

Using the computerized statistics package available to you, obtain a table showing the relationship between V04 (Presidential Vote) and V34 (Ideology) for the 1972 presidential election. In order to run this table, it will be necessary to select only certain cases from the dataset. Here we look only at data from the 1972 presidential election, thus selecting only those cases for which we have data from 1972. In SPSS, this can be done by using a *select if* statement. This statement selects cases according to the criterion that follows the "if" part of the statement. For this exercise, the statement would look like the following:

temporary.
select if (v01 eq 1).

Then the *crosstabs* command would be run. The statements above define a *temporary* selection of cases. This means that the selection is in effect *only* for the command that immediately follows the *select if* command. The second

part of the statement selects cases only if V01 (Year of Survey) is equal to a "1" (1972).¹

You should be able to determine which of these two is the column variable, which is the row variable, and how to lay out the table. Remember that in this book, we have adopted the convention to have the independent variable as the column variable and the dependent variable as the row variable. If you have also adopted this convention, your table should look like Table IV-1, below. You can check the table you generated against it to make sure that you properly requested the table.

The table has V34 at the top—as the independent variable—and V04 on the side—as the dependent variable—because it makes sense to think that ideology may affect the vote, not the other way around. Before going on with Exercise 1, make sure you understand Table IV-1 and the commands you used to generate it. By comparing the column entries and otherwise examining the numbers and percentages, you should be able to answer the following questions:

1. Overall, what percentage of respondents in the table voted for the Republican candidate, Richard Nixon? What percentage voted for the Democratic candidate, George McGovern?
2. How many respondents in the table were in each of the five categories of V34? Is this the kind of distribution you would have expected for this question?

Let us now look at the two-party vote by ideology. We might want to do this in order to convert V04 (Presidential Vote) from a three category nominal-level variable to a two category variable which can be treated as ordinal. In order to do this, we must recode V04 in the same manner that it was recoded in Chapter III (i.e., setting other party vote to missing data).² Using whatever recoding commands are available in the compute package you are using, perform this recode and then rerun the relationship between the recoded V04 (which we will refer to as V04r) and V34 (Ideology). Remember that you will need to run the temporary *select if* statement you developed above in order to obtain a table that only contains data from 1972. The correct table is reproduced below as Table IV-2. You can check the table you generated against it to make sure that you properly requested the table. By comparing the column entries and otherwise examining the numbers and percentages, you should be able to answer the following questions:

1. What percentage of the respondents who identified themselves as liberals voted for the Democratic candidate? How does this compare with the percentage voting Democratic in the other four categories? How would you describe the overall relationship between these two variables?

TABLE IV-1: 1972 PRESIDENTIAL VOTE BY IDEOLOGY

<i>Presidential Vote</i>	<i>Ideology</i>					
	<i>Liberal</i>	<i>Slightly Liberal</i>	<i>Middle of the Road</i>	<i>Slightly Conservative</i>	<i>Conservative</i>	
Democrat	82.2% (125)	55.6% (94)	30.8% (131)	15.2% (40)	9.6% (19)	33.9% (409)
Republican	17.1% (26)	43.2% (73)	68.2% (290)	83.3% (220)	88.3% (174)	64.9% (783)
Other	.7% (1)	1.2% (2)	.9% (4)	1.5% (4)	2.0% (4)	1.2% (15)
	100.0% (152)	100.0% (169)	100.0% (425)	100.0% (264)	100.0% (197)	N = 1207

Exercise 1B

Now that you have calculated the relationship between ideology and presidential vote for 1972, we can begin to utilize the pooled cross-sectional nature of this dataset for its full value. Let us examine the relationship between ideology and two-party presidential vote for each of the six presidential elections for which we have data. By looking at each year's data in turn, we can see whether the relationship between ideology and presidential vote substantially changes or whether it remains fairly consistent. We can also cast light on the question of whether the American public is becoming more ideological in its voting preferences.

If the American public were becoming more ideological, we would expect to see a stronger relationship between ideology and voting. An easy way to see this is to note whether the summary statistic we calculated for each table is increasing over time. At the bottom of Table IV-2 you can see the statistic tau-*c* that was calculated for the data in the table. The level of tau-*c* for this table (.49) is high—thus there is a strong relationship between ideology and presidential vote in 1972.

In order to address this question, you will need to run six three-way tables that display V04 (presidential vote)—the row variable—by V34 (ideology)—the column variable—by V01 (year of survey)—the layer variable. Three-way tables can be thought of as a stacked series of two-way tables. Each layer in the stack described here is a table that shows V04 by V34 for a given level of V01. There are six tables in this stack since there are six levels of V01—or six electoral years for which we have data. Adopting the same convention we have used before (the row variable being the dependent variable and the column variable being the independent variable), we can call the layer variable the control variable. The three-way table thus controls for electoral year, meaning that we are looking at the relationship between V04 and V34 within each electoral year.

In SPSS the commands to generate three-way tables look much like the commands to generate two-way tables. The *crosstabs* command to generate the six tables for this analysis is:

```
crosstabs tables=v04 by v34 by v01/cells  
counts col/statistics tau-c.
```

You should make sure that you thoroughly understand what this command does before continuing with these exercises.

After examining the six tables you generated, you should be able to answer the following questions:

1. Has tau-*c* increased substantially since 1972? Is the American public becoming more ideological in its presidential voting? Does ideology predict the presidential vote better in 1992 than it did in 1972?

To this point, we have been considering ideology as an independent variable which predicts presidential vote, the dependent variable. We can now look at ideology as a dependent variable and address the related question of ideological movement in the American public. The question of whether the American public is becoming more conservative is one that is often discussed in the media. We can examine this question by looking at the relationship between ideology and the year in which each survey was conducted. To do this, obtain a table showing the relationship between ideology (V34) and the year in which the data were collected (V01). You should be able to set up the table properly, including specifying which is the row and which the column variable, and telling SPSS or whatever statistical analysis package you are using to supply the proper percentages. You can check your results with those in Table IV-3.

In this case, we are interested in seeing whether there is a noticeable trend in the data. By looking at the table you obtained for V34 by V01 you should be able to answer the following questions:

1. Has the percentage of people who call themselves conservative increased significantly from 1972 to 1992? Has the percentage of people who call themselves liberal decreased significantly over the same period of time?
2. Are the results in Table IV-3 what you expected? How would you attempt to explain the findings?

TABLE IV-2: TWO-PARTY 1972 PRESIDENTIAL VOTE BY IDEOLOGY

Presidential Vote	Ideology					
	Liberal	Slightly Liberal	Middle of the Road	Slightly Conservative	Conservative	
Democrat	82.8% (125)	56.3% (94)	31.1% (131)	15.4% (40)	9.8% (19)	33.9% (409)
Republican	17.2% (26)	43.7% (73)	68.9% (290)	84.6% (220)	90.2% (174)	64.9% (783)
	100.0% (151)	100.0% (167)	100.0% (421)	100.0% (260)	100.0% (193)	N = 1192

Kendall's tau-c = .49

Exercise 1C

Related to the questions above is one concerning the meaning of party identification. We had a great deal to say about party identification (V04) in Chapter III and discussed in detail the coding of V04 and possible recoding schemes for V04. Now let us discuss what party identification means to the individual American citizen. Early voting studies (Campbell, Converse, Miller, and Stokes 1960) stressed the socialization aspects of party identification; children tended to show similar party identifications as their parents. This socialization argument and the strong tendency for party identification to be linked to presidential voting has been used to explain, for example, why conservative white southerners might continue to vote Democratic even though they would not to be in tune with the leadership of the national Democratic party. More recently, voting studies and the popular media have begun examining the ideological nature of party identification and asking the question of whether party identification can be predicted more closely from ideology now than it could in years past. Put another way, the question might be phrased as whether more ideological parties have resulted in the American public more rationally choosing its party affiliations.

Evidence that supports or refutes the notion that party identification has become more ideological can be gathered by obtaining a series of tables that look at the relationship between ideology (V34) and party identification (V09) over time (V01). Before obtaining these tables, it would most likely help to recode V04 in one of the two ways discussed in Chapter III. It would also help to collapse the number of categories in V34 to three—liberal, moderate, and conservative. These two recodes would provide a series of six three by three tables which can be easily compared over the 1972 through 1992 time period. Obtain these six tables and specify either tau-*b* or gamma as a statistic for each of the tables. By looking across the six tables, you should be able to answer the following questions:

1. What percentage of liberals call themselves Democrats in each of the six tables? Has this percentage increased perceptively from 1972 through 1992?
2. What percentage of conservatives call themselves Republicans in each of the six tables? Has this percentage increased perceptively from 1972 through 1992?
3. Likewise, what percentage of conservatives call themselves Democrats in each of the six tables? Has this percentage decreased? And what percentage of liberals call themselves Republicans in each of the six tables? Has this percentage decreased?
4. Has the value of the statistic (either tau-*b* or gamma) that you specified for each table increased monotonically from 1972 to 1992? What do you think explains the relationships that you have found?

Exercise 2: Party Identification and Presidential Voting

Our second exercise compares two presidential elections—1988 and 1992—and looks at the tendency for those who identify with a political party to vote for the candidate of that party, something that was discussed in Chapter II. This tendency has often been termed the loyalty rate, and its opposite (when strong party identifiers vote for the other party's candidate) has been termed the defection rate. The elections of 1988 and 1992 were particularly interesting in this regard since the same person, George Bush, was the Republican candidate in both years. In order to examine loyalty rates for these two elections, obtain tables for presidential vote (V04)—the three-party version of this—by party identification (V09) for the years 1988 and 1992.³ It should be apparent to you that Democratic and Republican loyalty rates changed rather dramatically from 1988 to 1992. In 1988, approximately 98% of strong Republicans voted for their party's candidate for president while slightly less than 93% of strong Democrats voted for their party's candidate. In 1992, these numbers were dramatically different; some 93% of strong Democrats voted for their party's candidate while 85.6% of strong Republicans voted for their party's candidate.

What are some reasons why strong party identifiers might defect to the opposite party? One might be a particularly attractive candidate nominated by the opposition party. This

TABLE IV-3: IDEOLOGY BY YEAR OF SURVEY

Ideology	Year of Survey						
	1972	1976	1980	1984	1988	1992	
Liberal	12.1% (188)	12.1% (231)	11.8% (118)	12.7% (197)	10.3% (147)	14.2% (259)	12.3% (1140)
Slightly Liberal	13.7% (212)	12.0% (220)	13.5% (136)	12.9% (201)	13.1% (186)	13.3% (243)	13.0% (1206)
Middle of the Road	37.4% (579)	37.7% (720)	30.6% (307)	33.4% (520)	31.3% (446)	31.5% (574)	34.0% (3146)
Slightly Conservative	20.8% (322)	18.5% (353)	21.0% (211)	20.1% (313)	21.7% (309)	20.3% (370)	20.3% (1877)
Conservative	16.0% (247)	19.7% (376)	23.1% (232)	20.8% (324)	23.6% (337)	20.6% (376)	20.4% (1892)
	100.0% (1548)	100.0% (1908)	100.0% (1004)	100.0% (1555)	100.0% (1425)	100.0% (1822)	N = 9261

dataset contains feeling thermometer scores for presidential candidates⁴ which could be run by party identification to see whether this is the case in any given presidential election. A second reason might be the presence of a particularly appealing third party candidate—Ross Perot in 1992. A third reason might be the state of the economy and how individuals feel about their own place in that economy.⁵ Given the changed economic conditions from 1988 to 1992, might we not expect that the changeover in the loyalty rate is tied up with perceptions of the economy? There are several variables in this dataset which could be to examine this possibility—V32 assesses individual perceptions of national economic conditions, and V33 asks about personal financial conditions.

By now you should be able to set up tables to examine the relationship in 1988 and 1992 between V09 (Party Identification) and V04 (Presidential Vote) controlling for either V32 or V33. This will require you to re-examine the exercises above to see how the three-way table commands were developed and how to use the *select if* command. Then you can adapt those commands to the current exercise. After you run the tables you are interested in, you might want to compare the results you obtained for 1988 and 1992 with the elections of 1976 and 1980, elections which were somewhat similar to the former ones. In 1980 many individuals were critical of national economic conditions and worried about their own personal financial conditions. As a presidential candidate, Ronald Reagan attacked President Jimmy Carter's record on handling the economy. Would you expect to see a similar changeover in loyalty rates from 1976 to 1980 as you did between 1988 and 1992? Would you think that the same variables that explain this changeover in 1988 and 1992 would also explain it in 1976 and 1980? You might also want to compare the elections of 1972 with 1976. In 1972, President Richard Nixon won a re-election of landslide proportions, but was forced to resign from office in 1974 as a result of the Watergate scandal. Would we think that Democrats (possibly

disenchanted with the candidacy of George McGovern) might disproportionately defect to vote for Nixon in 1972 but that Republicans (disgusted with the Watergate scandal) might disproportionately defect to vote for Carter in 1976?

Exercise 3: The Voting Behavior of Southern Whites

Let us now turn our attention to a three variable relationship in which the control variable is *not* the year in which the election took place. One of the more interesting areas of study in recent electoral analysis has been that of southern voting patterns, in particular the voting patterns of southern whites. Southern whites have had an historical affinity to the Democratic party but more and more of these voters have begun routinely to vote Republican both for president and in congressional elections.⁶ One of the problems in looking at the voting behavior of southern whites has been the small number of cases available in any given American National Election Study. The pooled cross-sectional nature of the current dataset allows us to look in detail at southern whites with the luxury of having a sufficient number of cases to make this analysis valuable.

In order to begin this analysis, we will need to create a variable that isolates individuals who are both southerners and white. Information on the region of the country in which an individual lives is contained in V69. Category number 3 of V69 is labeled "South." Information on an individual's race is contained in V59. Category number 1 is labeled "White." We can therefore create a new variable that is coded a "1" for individuals who are both southern and white (V69 is a "3" and V59 is a "1") and a "0" for everybody else. The commands to do this in SPSS follow:

```
compute whsouth=0.
if (v59 eq 1 and v69 eq 3) whsouth=1.
variable labels whsouth 'Southern Whites'.
value labels whsouth 0 'Other' 1 'Southern White'.
```

TABLE IV-4: PRESIDENTIAL VOTE (1972–1992) BY VIEW OF CHANGE IN CIVIL RIGHTS FOR WHITE SOUTHERNERS

Presidential Vote	Views of Change in Civil Rights			
	Too Slowly	About Right	Too Quickly	
Democrat	56.1% (50)	41.2% (322)	33.2% (279)	38.1% (651)
Republican	43.9% (39)	58.8% (459)	66.2% (561)	61.9% (1059)
	100.0% (89)	100.0% (780)	100.0% (840)	N = 1709

Gamma = .21

The intent of these commands should be apparent—to create a new variable that allows us to compare white southerners with all other respondents in the country. These commands first create a new variable (*whsouth*) that only has values of “0”. The new variable is then set to “1” for cases where V59 (Race) is equal to “1” (White) and V69 (Region) is equal to “3” (South). The new variable is then given a variable label and its values are also labeled. Of course, you could name the new variable anything you like; we use *whsouth* only as an example. Before working with your new variable, you should execute a *frequencies* command on the new variable—or crosstabulate it with Region (V69) and Race (V59)—in order to determine that your commands worked properly.

The dataset includes a question about how fast changes in civil rights are occurring (V44). The categories for this variable are: 1 - Too slowly; 2 - About right; 3 - Too fast. Let us obtain a table that looks at the relationship between this variable and presidential vote for southern whites versus all other voters. Because of the complicated way in which the two tables were created we have reproduced them below so that you can check your results.

After examining these two tables you should be able to answer the first of the following questions, and we hope you will be intrigued enough by this exercise to seek answers to those questions that follow the first one:

1. Is there a stronger relationship between opinions on how quickly changes in civil rights are occurring and the presidential vote for white southerners or for other respondents?
2. Why do you think that the strength of the relationship (as measured by gamma) between civil rights and the presidential vote would only be about half as strong for white southerners as it is for the remainder of the respondents? Put another way, why do you think that opinions on how quickly civil rights is changing predict presidential vote worse for white southerners than those opinions predict presidential vote for other respondents?
3. How would you think that the relationships you identified above would vary by the age of the white south-

erners in this sample? Try running presidential vote (V04) by civil rights change (V44) by age (V60) to see if age has any effect here.⁷

4. What other variables might you want to use to try to predict white southerners’ presidential votes? Do you think that these relationships would be similar or different for elections in the early part of this series (the 1970s) and the later part of this series (late 1980s and early 1990s)? Obtain the relevant tables to examine these factors.

Exercise Four: The Gender Gap

Our final exercise focuses on an even more complex set of relationships and is intended to stimulate your thinking for further work. It is also intended to illustrate further reasons for examining multivariate tables.

Consider the possible existence of a “gender gap” in support for Democratic and Republican presidential candidates by men and women. Is there really a gender gap in presidential candidates’ support? If so, why? When did it develop? Has it grown in proportions over the time period for which we have data?

Obtain a table which presents a crosstabulation of the two-party presidential vote (V04) by respondent’s gender for the pooled cross-sectional dataset as a whole. You should already know how to do this. This table will show that for the sample as a whole, women have tended to vote for the Democratic candidate for president over the Republican candidate by about five percentage points. Two questions should suggest themselves at this point:

1. Has this gender gap been fairly constant over the time period or did it grow over time?
2. What variables help us explain and understand the differing voting patterns of men and women?

Let us take these questions one at a time. You can easily obtain the six tables necessary to address the first question. You have already obtained tables in earlier exercises that look at the relationship between two variables controlling for the electoral year. Obtaining these tables will be very similar. Once you do this, what do you see? Has the gender gap remained constant, has it grown, has it lessened, or is it apparent for some years and not for others?

To answer the second question, we will need to run a variety of tables that look at presidential vote by gender controlling for other variables. Some possible control variables are suggested here.

One possible explanation for a gender gap is that men and women have different sets of opinions on public policy issues which more directly affect women than men. The dataset included with this module contains a number of questions which might be more relevant to women than to men. Some of these are the question on abortion (V45); the women’s role scale (V46); and the guaranteed job scale (V41).

You should be able to run the necessary tables to examine the relationship between gender, these public policy issues, and the vote. Determine which additional variables need to be introduced into the analysis and obtain the proper

TABLE IV-5: PRESIDENTIAL VOTE (1972–1992) BY VIEW OF CHANGE IN CIVIL RIGHTS FOR ALL OTHER RESPONDENTS

Presidential Vote	Views of Change in Civil Rights			
	Too Slowly	About Right	Too Quickly	
Democrat	80.2% (601)	49.4% (1422)	35.8% (596)	49.5% (2619)
Republican	19.8% (149)	50.6% (1458)	64.2% (1067)	50.5% (2673)
	100.0% (750)	100.0% (2880)	100.0% (1663)	N = 5293

Gamma = .44

tables. A good place to start would be to obtain tables for whatever public policy issues you are interested in by gender. The table for opinions on abortion (V45) by gender (V58) is included as Table IV-6 in order to get you started. Are women's opinions on abortion very different from men's? Are women's opinions on abortion what you expected them to be?⁸

The tables you obtain will tell you if there are, indeed, gender differences on these issues we assume are more relevant to women than to men. After looking at these tables, additional multivariate tables will need to be generated and analyzed. You might want to look at how feeling thermometer scores for the candidates differ between male and female respondents; do female respondents "like" Democratic candidates better than they "like" Republican candidates? Analyze the results of these tables and see what conclusions you draw concerning the nature and sources of the gender gap.

Testing Your Own Theory

One way of studying voting behavior might be to have the computer generate all possible tables that could be constructed from these data. Such a strategy is neither practical or useful. Tables by themselves tell you very little. They have to be interpreted and explained. This usually is done in terms of some theory or generalization about how people behave.

To make research efforts more meaningful, and to avoid wasted effort, we need to frame the question we are studying carefully and to be sure that we have some sound reason for studying it. That reason usually is directly related to some theoretical concepts or ideas that we have learned from the work of others. The general rule is to apply what we already know to discover something new.

It is important to explicate why we are looking at a particular set of relationships, because that justification becomes part of the explanation of what the contingency tables tell us. Political scientists are far more interested in

TABLE IV-6: OPINIONS ON ABORTION BY GENDER (1972–1992)

Opinions on Abortion	Gender		
	Male	Female	
Never Permit	10.6% (594)	12.5% (930)	11.7% (1524)
For Rape, etc.	35.1% (1974)	37.8% (2803)	36.7% (4777)
If Need Established	18.8% (1059)	16.4% (1215)	17.4% (2274)
Always Permit	35.5% (1994)	33.2% (2463)	34.2% (4457)
	100.0% (5621)	100.0% (7411)	N = 13,032

the relationship among variables than they are in the actual percentages of some group that voted one way or another, as these relationships help to explain behavior. We especially want to know why the variables are related the way that they are. We also want to know why two variables that our theory predicts *should* be related might wind up not to be when we look at the data. Often, the lack of a find or the fact that two variables are unrelated is just as important as finding that two variables are strongly related, but only when we have a pre-existing theory or hypothesis that predicts that the two variables should be related.

Most research is hypothesis testing. We start from a theory developed by reading other people's work; we generate hypotheses from that theory; and we test those hypotheses by comparing our predicted relationships to those resulting from data analysis. If the data bear out our hypotheses, we can claim some support for the theory; if the data do not bear out our hypotheses, we can conclude that the theory is not supported. In either event, the conclusions are valuable.

Using what you already know about voting behavior, as well as your intuition, design a short research project that examines some aspect of voting behavior. Specifically, you should: (1) formulate a question about the relationship among a set of variables; (2) justify the selection of variables and the expectations you have about the connection between the variables; (3) test your ideas by obtaining the necessary tables; and (4) carefully interpret the tables and write up your conclusions. In considering the analyses you want to do to examine the relationships you identify, think about whether you will need to look at those relationships in one particular year, to look at relationships between variables and the electoral year, or to look at relationships in the pooled data set for either several years or for the entire time period.

Notes

1. If you are using a statistical package other than SPSS, your commands to subset the dataset will, of course, look different from these.
2. For the rest of these exercises, when we use presidential vote as a dependent variable, we will only use the two party vote. This simplifies the exercises. In each case, however, you should consider how adding John Anderson's vote in 1980 or Ross Perot's vote in 1992 would change (or complicate) the relationship between the independent and dependent variables.
3. Refamiliarize yourself with the use of the *select if* statement that we used in Exercise 1a, above. For this exercise, we will be selecting cases that show either a "5" (1988) or a "6" (1992) on V01.
4. The Democratic candidate's feeling thermometer score is V25 while the Republican candidate's feeling thermometer is V27.
5. Of course, all three of these reasons interact. Individuals who are dissatisfied with the state of the economy might view an opposition candidate as more attractive or might find a third party candidate more appealing than their own party's candidate.
6. The results of the congressional elections of 1994, for example, show that there are more southern Republican members of Congress than there are Democratic for the first time since Reconstruction.
7. You will need to recode age (V60) into a smaller number of categories to make your analysis easier to interpret.
8. You might note the large number of respondents included in this table. Since the table simply runs abortion against gender, no distinction is made between people who have voted and those who have not. As a second step in analyzing this relationship, you should probably run tables that look at abortion opinions by gender for voters and non-voters separately in order to see whether voters' opinions are similar to non-voters' opinions.

Chapter V

User Information

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 dataset 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 dataset, but in order to simplify the dataset, 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 dataset 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

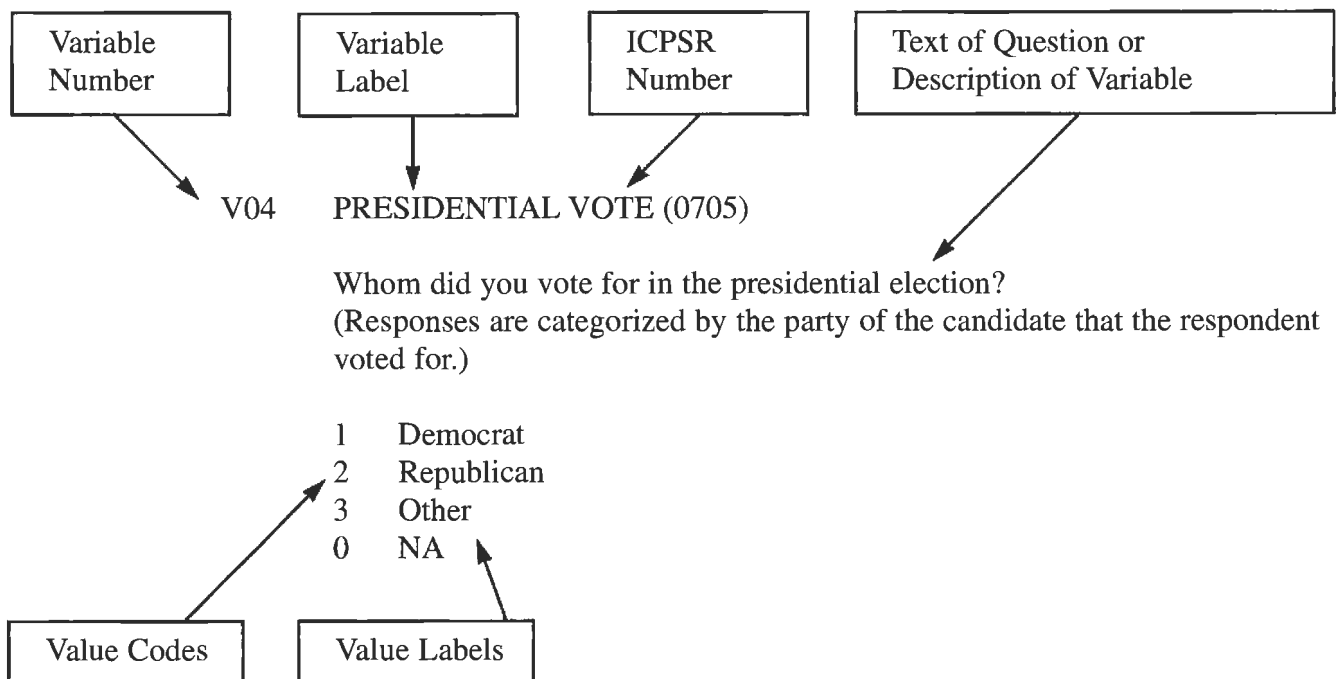
The codebook (Chapter VI) provides both a description of the data and information necessary for using the data. Every available variable is listed in the codebook, and the entry for each variable contains the information needed to use the variable in the data analysis. Below is a sample

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.

Explanation of Codebook Information

1. *Variable number.* Each variable in the dataset 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 dataset.
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.* This is the variable number in the codebook for the ANES Cumulative Data File (ICPSR study number 8475). If there is a need to refer to the original version of the question or variable for one of the items in the dataset, the ICPSR reference number will help locate that information. The variables in this module generally are simplified or condensed to make the analysis easier for the student. In a few cases, there are two numbers listed, indicating that two differ-

Sample Codebook Entry



ent versions of the variable were used during this time period and were combined into one variable for this dataset.

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 anal-

ysis 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 “0” 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.”

Chapter VI

Codebook

WT1 WEIGHT FACTOR 1 (0009)

See the appendix for a discussion of weight factor 1.

WT2 WEIGHT FACTOR 2

See the appendix for a discussion of weight factor 2.

V01 YEAR OF SURVEY (0004)

Year in which the survey was conducted.

- 1 1972
- 2 1976
- 3 1980
- 4 1984
- 5 1988
- 6 1992

A. VOTING BEHAVIOR AND RELATED ITEMS

V02 REGISTERED TO VOTE? (0737)

Were you registered to vote?
(Not available for 1972 and 1976.)

- 1 Yes
- 2 No
- 0 NA

V03 VOTED IN ELECTION (0702)

Did you vote in the presidential election?

- 1 Voted
- 2 Did not vote
- 0 NA

V04 PRESIDENTIAL VOTE (0705)

Whom did you vote for in the presidential election?
(Responses are categorized by the party of the candidate that the respondent voted for.)

- 1 Democrat
- 2 Republican
- 3 Other
- 0 NA

V05 TIME OF VOTE DECISION (0712)

How long before the election did you decide how you were going to vote?

- 1 All along
- 2 When candidate announced
- 3 During convention
- 4 Post-convention period
- 5 Last two weeks of campaign
- 6 Election day
- 0 NA

V06 SENATE VOTE (0708)

Whom did you vote for in the U.S. Senate election?
(Responses are categorized by the party of the candidate that the respondent voted for.)

- 1 Democrat
- 2 Republican
- 0 NA

V07 HOUSE VOTE (0707)

Whom did you vote for in the House of Representatives election? (Responses are categorized by the party of the candidate that the respondent voted for.)

- 1 Democrat
- 2 Republican
- 0 NA

V08 HOUSE RACE TYPE (0902)

Classification of the type of race for the U.S. House of Representatives in the respondent's district. Districts were classified as to whether: (a) a Democratic incumbent was running for reelection; (b) no incumbent was running for reelection; or (c) a Republican incumbent was running for reelection.

(Not available for 1972.)

- 1 Democratic incumbent
- 2 No incumbent
- 3 Republican incumbent
- 0 NA

V09 PARTY IDENTIFICATION (0301)

Generally speaking, do you think of yourself as a Republican, a Democrat, an independent, or what? If Democrat or Republican, would you call yourself a strong or weak Democrat or Republican? If independent, are you closer to the Republican or the Democratic Party?

- 1 Strong Democrat
- 2 Weak Democrat
- 3 Independent Democrat
- 4 Independent
- 5 Independent Republican
- 6 Weak Republican
- 7 Strong Republican
- 0 NA

B. POLITICAL INVOLVEMENT ITEMS

V10 INTEREST IN ELECTION (0310)

How interested have you been in the political campaigns this year?

- 1 Not much
- 2 Somewhat
- 3 Very much
- 0 NA

V11 WATCH TV NEWS ABOUT CAMPAIGN? (0724)

Did you watch any programs about the campaign on television?

(Not available for 1988.)

- 1 No
- 2 Yes
- 0 NA

V12 READ NEWSPAPERS ABOUT CAMPAIGN? (0727)

Did you read about the campaign in any newspaper?

- 1 No
- 2 Yes
- 0 NA

V13 MEDIA EXPOSURE INDEX (0728)

What sources of media did you use to follow the campaign? (Respondents were categorized according to the number of the following sources used: newspaper, television, radio, and magazines.)

(Not available for 1988.)

- 1 No media
- 2 One media source
- 3 Two media sources
- 4 Three media sources
- 5 All four media sources
- 0 NA

V14 ATTEND POLITICAL MEETINGS? (0718)

Did you go to any political meetings, rallies, dinners, or things like that in support of a particular candidate?

- 1 No
- 2 Yes
- 0 NA

V15 WORK FOR PARTY OR CANDIDATE? (0719)

Did you do any work for one of the parties or candidates other than attending meetings or rallies?

- 1 No
- 2 Yes
- 0 NA

V16 DONATE MONEY? (0721)

Did you give money to an individual candidate or to a political party this election year?

- 1 No
- 2 Yes
- 0 NA

C. CANDIDATE IMAGE ITEMS

V17 DEMOCRATIC CANDIDATE INSPIRING? (0353)

Does the phrase “inspiring” describe the Democratic presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V18 DEMOCRATIC CANDIDATE KNOWLEDGEABLE? (0354)

Does the phrase “knowledgeable” describe the Democratic presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V19 DEMOCRATIC CANDIDATE MORAL? (0355)

Does the phrase “moral” describe the Democratic presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V20 DEMOCRATIC CANDIDATE STRONG LEADER? (0356)

Does the phrase “strong leader” describe the Democratic presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V21 REPUBLICAN CANDIDATE INSPIRING? (0365)

Does the phrase “inspiring” describe the Republican presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V22 REPUBLICAN CANDIDATE KNOWLEDGEABLE? (0366)

Does the phrase “knowledgeable” describe the Republican presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V23 REPUBLICAN CANDIDATE MORAL? (0367)

Does the phrase “moral” describe the Republican presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V24 REPUBLICAN CANDIDATE STRONG LEADER? (0368)

Does the phrase “strong leader” describe the Republican presidential candidate extremely well, quite well, or not well at all?

(Not available for 1972 and 1976.)

- 1 Extremely well
- 2 Quite well
- 3 Not well
- 0 NA

V25 DEMOCRATIC CANDIDATE THERMOMETER SCORE (0424)

On a scale of 0 degrees to 100 degrees, how warm or cool do you feel toward the Democratic presidential candidate? (Respondents are categorized into the five groups listed below.)

- 1 0-20 degrees
- 2 21-40 degrees
- 3 41-60 degrees
- 4 61-80 degrees
- 5 81-100 degrees
- 0 NA

V26 DEMOCRATIC VP CAND THERMOMETER SCORE (0425)

On a scale of 0 degrees to 100 degrees, how warm or cool do you feel toward the Democratic vice presidential candidate? (Respondents are categorized into the five groups listed below.)

- 1 0-20 degrees
- 2 21-40 degrees
- 3 41-60 degrees
- 4 61-80 degrees
- 5 81-100 degrees
- 0 NA

V27 REPUBLICAN CANDIDATE THERMOMETER SCORE (0426)

On a scale of 0 degrees to 100 degrees, how warm or cool do you feel toward the Republican presidential candidate? (Respondents are categorized into the five groups listed below.)

- 1 0-20 degrees
- 2 21-40 degrees
- 3 41-60 degrees
- 4 61-80 degrees
- 5 81-100 degrees
- 0 NA

V28 REPUBLICAN VP CAND THERMOMETER SCORE (0427)

On a scale of 0 degrees to 100 degrees, how warm or cool do you feel toward the Republican vice presidential candidate? (Respondents are categorized into the five groups listed below.)

- 1 0-20 degrees
- 2 21-40 degrees
- 3 41-60 degrees
- 4 61-80 degrees
- 5 81-100 degrees
- 0 NA

D. GOVERNMENT PERFORMANCE ITEMS

V29 PRESIDENTIAL JOB APPROVAL (0450)

Do you approve or disapprove of the way that the President is handling his job?

- 1 Approve
- 2 Disapprove
- 0 NA

36 Codebook

V30 MOST IMPORTANT NATIONAL PROBLEM (0875)

What do you think is the single most important problem facing this country? (Responses have been categorized into the five groups listed below.)

- 1 Economy
- 2 Foreign affairs
- 3 Social welfare
- 4 Crime, public order
- 5 Other
- 0 NA

V31 BEST PARTY FOR PROBLEM (9012)

Which political party do you think would be most likely to get the government to do a better job in dealing with this problem—the Republicans, the Democrats, or wouldn't there be much difference between them?

- 1 Democrats
- 2 No difference
- 3 Republicans
- 0 NA

V32 NATIONAL ECONOMIC CONDITIONS (0871)

Would you say that over the past year the nation's economy has gotten better, stayed the same, or gotten worse?

(Not available for 1972 and 1976.)

- 1 Better
- 2 Same
- 3 Worse
- 0 NA

V33 PERSONAL FINANCIAL SITUATION (0880)

Would you say that you and your family are better off, the same, or worse off financially than you were a year ago?

- 1 Better
- 2 Same
- 3 Worse
- 0 NA

E. IDEOLOGICAL ITEMS

V34 IDEOLOGY (0803)

Respondent's self-placement on a scale running from "Liberal" to "Conservative."

- 1 Liberal
- 2 Slightly liberal
- 3 Moderate
- 4 Slightly conservative
- 5 Conservative
- 0 NA

V35 DEMOCRATIC CANDIDATE: IDEOLOGY (9088)

Respondent's placement of the Democratic presidential candidate on a scale running from "Liberal" to "Conservative."

- 1 Liberal
- 2 Slightly liberal
- 3 Moderate
- 4 Slightly conservative
- 5 Conservative
- 0 NA

V36 REPUBLICAN CANDIDATE: IDEOLOGY (9096)

Respondent's placement of the Republican presidential candidate on a scale running from "Liberal" to "Conservative."

- 1 Liberal
- 2 Slightly liberal
- 3 Moderate
- 4 Slightly conservative
- 5 Conservative
- 0 NA

V37 DEMOCRATIC PARTY: IDEOLOGY (0503)

Respondent's placement of the Democratic Party on a scale running from "Liberal" to "Conservative."

- 1 Liberal
- 2 Slightly liberal
- 3 Moderate
- 4 Slightly conservative
- 5 Conservative
- 0 NA

V38 REPUBLICAN PARTY: IDEOLOGY (0504)

Respondent's placement of the Republican Party on a scale running from "Liberal" to "Conservative."

- 1 Liberal
- 2 Slightly liberal
- 3 Moderate
- 4 Slightly conservative
- 5 Conservative
- 0 NA

V39 PARTY IDEOLOGICAL DISTANCE (0556)

Perceived ideological difference between the Democratic and Republican Parties, based on the respondent's placement of the parties on the above ideological scales.

- 1 No difference
- 2
- 3
- 4
- 5 Great difference
- 0 NA

V40 IMPORTANT PARTY DIFFERENCES (0501)

Do you think there are any important differences in what the Republicans and Democrats stand for?

- 1 No
- 2 Yes
- 0 NA

F. ATTITUDES ON SPECIFIC ISSUES

V41 GUARANTEED JOB SCALE (0809)

Respondent's self-placement on a scale running from "Government should see to it that every person has a job and a good standard of living" to "Government should let each person get ahead on their own."

- 1 Guaranteed job
- 2
- 3
- 4
- 5 Individual on own
- 0 NA

V42 HEALTH PLAN SCALE (0806)

Respondent's self-placement on a scale running from "There should be a government insurance plan which would cover all medical and hospital expenses for everyone" to "All medical expenses should be paid by individuals or private insurance plans."

(Not available for 1980.)

- 1 Government plan
- 2
- 3
- 4
- 5 Private plan
- 0 NA

V43 MINORITY AID SCALE (0830)

Respondent's self-placement on a scale running from "Government should make every effort to improve the social and economic position of blacks" to "Government should not make any special effort to help blacks because they should help themselves."

- 1 Help minorities
- 2
- 3
- 4
- 5 Minorities on own
- 0 NA

V44 CIVIL RIGHTS CHANGE (0814)

Do you think the civil rights leaders are trying to push too fast, too slowly, or are they moving at about the right speed?

- 1 Too slowly
- 2 About right
- 3 Too fast
- 0 NA

V45 ABORTION (0837, 0838)

Respondent's agreement with one of the following statements: (1) By law, abortion should never be permitted; (2) The law should permit abortion only in case of rape, incest, or when the woman's life is in danger; (3) The law should permit abortion for reasons other than rape, incest, or danger to woman's life, but only after the need for the abortion has been clearly established; (4) By law, the woman should always be able to obtain an abortion as a matter of personal choice. (Note: The question wording for the second and third categories was slightly different in 1972 and 1976.)

- 1 Never permit
- 2 For rape, etc.
- 3 If need established
- 4 Always permit
- 0 NA

V46 ROLE OF WOMEN SCALE (0834)

Respondent's self-placement on a scale running from "Women should have an equal role with men in running business, industry and government" to "Women's place is in the home."

- 1 Equal role
- 2
- 3
- 4
- 5 In the home
- 0 NA

V47 DEFENSE SPENDING (0843)

Respondent's self-placement on a scale running from "We should spend much less on defense" to "We should greatly increase defense spending."

(Not available for 1972 and 1976.)

- 1 Greatly decrease
- 2
- 3
- 4
- 5 Greatly increase
- 0 NA

V48 U.S. WORLD INVOLVEMENT (0823)

Do you agree or disagree with this statement, "This country would be better off if we just stayed home and did not concern ourselves with problems in other parts of the world?"

- 1 Agree
- 2 Disagree
- 0 NA

V49 BEST PARTY TO AVOID WAR (0522)

Do you think the problem of keeping out war would be handled better in the next four years by the Republicans, by the Democrats, or about the same by both? (Not available for 1976.)

- 1 Democrats
- 2 No difference
- 3 Republicans
- 0 NA

G. GENERAL ORIENTATIONS TOWARD GOVERNMENT

V50 FEDERAL GOVERNMENT STRENGTH (0829)

Do you think the government in Washington is getting too strong or not strong enough for the good of the country and the individual person?

- 1 Not too strong
- 2 Too strong
- 0 NA

V51 TRUST IN GOVERNMENT (0604)

How much of the time do you think you can trust the government in Washington to do what is right?

- 1 Some of the time
- 2 Most of the time
- 0 NA

V52 GOVERNMENT RUN FOR ALL (0605)

Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?

- 1 Few big interests
- 2 Benefit of all
- 0 NA

V53 OFFICIALS DO NOT CARE (0609)

Do you agree or disagree with the statement, "I don't think public officials care much what people like me think?"

- 1 Agree
- 2 Neutral
- 3 Disagree
- 0 NA

V54 GOVERNMENT TAX WASTE (0606)

Do you think that people in the government waste a lot of money we pay in taxes or waste just some or very little of it?

- 1 A lot
- 2 Some or little
- 0 NA

V55 PEOPLE HAVE NO SAY (0613)

Do you agree or disagree with the statement, "People like me don't have any say about what the government does?"

- 1 Agree
- 2 Neutral
- 3 Disagree
- 0 NA

V56 POLITICS TOO COMPLICATED (0614)

Do you agree or disagree with the statement, "Sometimes politics and government seem so complicated that a person like me can't really understand what is going on?"

- 1 Agree
- 2 Neutral
- 3 Disagree
- 0 NA

V57 GOVERNMENT PAYS ATTENTION (0622)

How much attention do you feel the government pays to what the people think when it decides what to do — not much, some, or a good deal?

- 1 Not much
- 2 Some
- 3 A good deal
- 0 NA

H. SOCIAL AND DEMOGRAPHIC CHARACTERISTICS

V58 GENDER (0104)

Respondent's gender.

- 1 Male
- 2 Female
- 0 NA

V59 RACE (0106)

Respondent's race.

- 1 White
- 2 Black
- 3 Other
- 0 NA

V60 AGE (0102)

Respondent's age.

- 1 17-24
- 2 25-34
- 3 35-44
- 4 45-54
- 5 55-64
- 6 65-99
- 0 NA

V61 MARITAL STATUS (0147)

Respondent's marital status.

- 1 Married
- 2 Single
- 3 Divorced/separated
- 4 Widowed
- 0 NA

V62 EDUCATION (0140)

What is the highest grade of school you have completed?

- 1 Not a high school graduate
- 2 High school graduate only
- 3 Some college
- 4 College graduate
- 0 NA

V63 EMPLOYMENT STATUS (0116)

Are you employed, looking for work, retired, or what?

- 1 Employed
- 2 Unemployed
- 3 Retired/disabled
- 4 Homemaker
- 5 Student
- 0 NA

V64 OCCUPATION (0157)

What is your primary occupation? (Responses have been categorized into the four occupational groups listed below.)

- 1 Professional/managerial
- 2 Clerical/sales
- 3 Manual work
- 4 Farm
- 0 NA

V65 INCOME (0114)

What was the total income of your family last year? (Responses have been categorized into percentile groups. For example, someone in the fifth category had a family income that was in the top five percent of all family incomes in the country for that year.)

- 1 0-16 pctl
- 2 17-33 pctl
- 3 34-67 pctl
- 4 68-95 pctl
- 5 96-100 pctl

V66 SOCIAL CLASS (0148)

What would you call yourself: middle class or working class?

- 1 Working class
- 2 Middle class
- 0 NA

V67 RELIGION (0129, 0152)

What is your religious denomination? (Note: "Catholic" includes Eastern Orthodox churches; "Other Christian" includes Mormons, Quakers and Unitarians and others; "Other nonchristian" includes Muslims, Buddhists, and others.)

- 1 Protestant
- 2 Catholic
- 3 Other Christian
- 4 Jewish
- 5 Other nonchristian
- 6 None
- 0 NA

V68 CHURCH ATTENDANCE (0130)

How often do you go to religious services?

- 1 Every week
- 2 Almost every week
- 3 1-2 times a month
- 4 Few times a year
- 5 Never
- 0 NA

V69 REGION (0112)

Region of the country that the respondent lives in.

- 1 Northeast
- 2 North central
- 3 South
- 4 West

V70 URBANISM (0111)

Type of community that the respondent lives in.

- 1 Central city
- 2 Suburban area
- 3 Rural, small town

References

- Abramson, Paul R., John H. Aldrich, and David W. Rohde. 1982. *Change and Continuity in the 1980 Elections*. Washington, DC: Congressional Quarterly Press.
- Abramson, Paul R., John H. Aldrich, and David W. Rohde. 1987. *Change and Continuity in the 1984 Elections*. Revised edition. Washington, DC: Congressional Quarterly Press.
- Abramson, Paul R., John H. Aldrich, and David W. Rohde. 1991. *Change and Continuity in the 1988 Elections*. Revised edition. Washington, DC: Congressional Quarterly Press.
- Abramson, Paul R., John H. Aldrich, and David W. Rohde. 1995. *Change and Continuity in the 1992 Elections*. Revised edition. Washington, DC: Congressional Quarterly Press.
- Asher, Herbert B. 1992. *Presidential Elections and American Politics*. Pacific Grove, CA: Brooks/Cole Publishing Co.
- Black, Earl, and Merle Black. 1992. *The Vital South*. Cambridge: Harvard University Press.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald R. Stokes. 1960. *The American Voter*. New York: John Wiley and Sons.
- Converse, Philip. 1964. "The Nature of Belief Systems in Mass Publics," in *Ideology and Discontent*, edited by David E. Apter. New York: The Free Press, 1964.
- Converse, Philip. 1970. "Attitudes and Nonattitudes: The Continuation of a Dialogue," in *The Quantitative Analysis of Social Problems*, edited by Edward Tufte. Reading, MA: Addison-Wesley.
- Converse, Philip and Gregory Markus. 1979. "la Plus Ca Change . . . : The New CPS Election Study Panel," *American Political Science Review* 73 (March), 32-49.
- Edsall, Thomas B., and Mary D. Edsall. 1991. *Change Reaction*. New York: W.W. Norton.
- Feigert, Frank B. 1993. "The Ross Perot Candidacy and Its Significance." *America's Choice: The Election of 1992*. William Crotty, editor. Guilford, CT: Dushkin.
- Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven, CT: Yale University Press.
- Groves, Robert. 1989. "Actors and Questions in Telephone and Personal Interview Surveys," in *Survey Research Methods: A Reader*, edited by Eleanor Singer and Stanley Presser. Chicago: University of Chicago Press.
- Key, V. O., Jr. 1949. *Southern Politics in State and Nation*. New York: Alfred A. Knopf.
- Key, V. O., Jr. 1966 *The Responsible Electorate*. Cambridge: Harvard University Press.
- Kiewiet, D. Roderick. 1983. *Macroeconomics and Micropolitics*. Chicago: University of Chicago Press.
- Kish, Leslie. 1949. "A Procedure for Objective Respondent Selection within the Household," *Journal of the American Statistical Association*, 44, pp. 380-387.
- Ladd, Everett Carl Jr., and Charles D. Hadley. 1978. *Transformations of the American Party System*. Revised edition. New York: W.W. Norton.
- Lamis, Alexander P. *The Two-Party South*. New York: Oxford University Press.
- Lewis-Beck, Michael S., and Tom W. Rice. 1992. *Forecasting Elections*. Washington, DC: Congressional Quarterly Press.
- Luttbeg, Norman R., and Michael M. Gant. 1995. *American Electoral Behavior, 1952-1992*. Itasca, IL: F.E. Peacock Publishers.
- Miller, Arthur H., Warren E. Miller, Alden S. Raine, and Thad A. Brown. 1976. "A Majority Party in Disarray: Policy Polarization in the 1972 Election." *American Political Science Review* 70:753-778.
- Mueller, Carol M. 1988. *The Politics of the Gender Gap*. Newbury Park, CA: Sage Publications.
- Nie, Norman H., Sidney Verba, and John R. Petrocik. 1976. *The Changing American Voter*. Cambridge: Harvard University Press.
- Page, Benjamin I. 1978. *Choices and Echoes in Presidential Elections*. Chicago: University of Chicago Press.
- Pomper, Gerald M., et al. 1977. *The Election of 1976*. New York: David McKay.
- Pomper, Gerald M., et al. 1981. *The Election of 1980*. Chatham, NJ: Chatham House.
- Pomper, Gerald M., et al. 1985. *The Election of 1984*. Chatham, NJ: Chatham House.
- Pomper, Gerald M., et al. 1989. *The Election of 1988*. Chatham, NJ: Chatham House.
- Pomper, Gerald M., et al. 1993. *The Election of 1992*. Chatham, NJ: Chatham House.
- Quirk, Paul J., and Jon K. Dalager. 1993. "The Election: A 'New Democrat' and a New Kind of Presidential Campaign." *The Elections of 1992*. Michael Nelson, editor. Washington, DC: Congressional Quarterly Press.
- Rogers, Theresa. 1989. "Interviews by Telephone and in Person: Quality of Response and Field Performance," in *Survey Research Methods: A Reader*, edited by Eleanor Singer and Stanley Presser. Chicago: University of Chicago Press.
- Schneider, William. 1992. "The Suburban Century Begins." *The Atlantic* (July), 33-44.
- Sundquist, James. 1983. *The Dynamics of the Party System*. Washington, DC: The Brookings Institution.
- Wald, Kenneth D. 1987. *Religion and Politics in the United States*. New York: St. Martin's Press.
- Wattenberg, Martin. 1994. *The Decline of American Political Parties 1952-1992*. Cambridge: Harvard University Press.
- Weisberg, Herbert. 1987. "The Demographics of a New Voting Gap: Marital Differences in American Voting." *Public Opinion Quarterly* 51:335-343.
- Weisberg, Herbert, Jon Krosnick, and Bruce Bowen. 1989. *An Introduction to Survey Research and Data Analysis*, 2nd edition. Glenview, IL: Scott, Foresman, and Company.
- White, John K. 1993. "The General Election Campaign: Issues and Themes." *America's Choice: The Election of 1992*. William Crotty, editor. Guilford, CT: Dushkin.

Appendix: Technical Note

The data for this SETUPS are drawn from the 1972 through 1992 American National Election Study (ANES) Pre- and Post-Election Surveys. These data were collected by the Center for Political Studies (CPS). Since 1980 data collection has been funded by a grant from the National Science Foundation. The data are distributed by the Inter-university Consortium for Political and Social Research (ICPSR). All of the respondents interviewed in the six presidential elections held from 1972 to 1992 are included in this dataset. A total of 70 variables have been drawn, or in some cases created, from the data in the ANES Cumulative Data File (ICPSR study number 8475). Most of the variables have been recoded to simplify matters for the students by reducing the number of response categories, and in some cases two variables have been merged into a single item. The SETUPS codebook gives the original CPS/ICPSR reference number (or numbers, if more than one original variable was used) for each variable, so that you can refer to the full ANES codebook for the cumulative data file to obtain information on the original coding or the exact wording of each item.

The design of the six ANES studies varied somewhat. The sampling frames for 1972, 1980, 1984, and 1984 are similar; the data for these years are unweighted. In 1976, the data are weighted to correct for respondents who were part of the panel study conducted in the 1972-76 period. The weights calculated for those years cause the 1976 sample to increase in size from 2248 to 2870 when one weights the data. The 1992 ANES sample is more complex than those used in previous years. Some of the respondents were part of a panel study that originated in 1990; others were part of a freshly-drawn cross-sectional sample for 1992. Because of differences in the response rate between the two portions of the sample, and because of the unequal probability of selection of respondents in different size households, the data should be weighted for any analysis. The 1992 weights keep the weighted and unweighted samples approximately equal in size; there are 2255 unweighted respondents and 2252 weighted respondents. The appropriate weight (variable 0009 in the ANES 1952-1992 Cumulative file) for these years is included in the dataset as WT1. The weighting of the data generally will be transparent to the student users.

In addition to the weight variable calculated by the ANES staff (WT1), we have calculated a second weight variable (WT2) for those who might be concerned about the unequal number of respondents in each electoral sample. Note 7 in Chapter 3 describes this situation in detail. For those instructors who are concerned that the percentage of respondents in any given electoral sample ranges from a low of 11.6% in 1980 to a high of 20.5% in 1976, we suggest that you weight the data by variable WT2. This sets each sample equal to 1/6 the size of the dataset as a whole. The WT2 weight incorporates the WT1 weight in addition to adjusting for the unequal sample sizes across the years. The tables in this monograph are calculated on the data

weighted by WT1; if you *do* weight the data by WT2, you will need to recalculate all of the tables in the monograph.

The dataset for this SETUPS is available from the ICPSR as study 6572. The data may be ordered either on: (a) magnetic tape, suitable for use with mainframe or super-mini computers; or (b) 3.5" diskettes, suitable for use with MS-DOS machines. The version that you want will be shipped by the ICPSR when this SETUPS is ordered from the APSA. Forms for the technical specifications will be sent when you place your order.

This SETUPS is prepared especially for SPSS (or SPSS for Windows) and SAS (or SAS-PC) users, but those who prefer to use a different statistical package should be accommodated. Four files will be shipped on the tape or diskette sent by the ICPSR: (a) an SPSS export file; (b) the raw data; (c) the SPSS file definition statements for constructing an SPSS system file; and (d) the SAS control statements for constructing a SAS file. Most SPSS users should be able to work with the export file and will not need the other files. However, it also is possible to construct an SPSS system file from the raw data file and the file definition statements, and this should work in any computing environment. If you choose this alternative, the file definition statements will have to be edited slightly by adding any necessary local job control statements at the beginning and by modifying the "data list" and "save outfile" statement to include the actual file specifications appropriate for the computing environment upon which SPSS is being run. The remainder of the file should not require editing.

SAS users can construct a SAS file from the raw data file and the file with SAS control statements, with some slight editing of the control statements to conform to the local computing environment. In some environments it also may be possible to import the SPSS export file directly into SAS.

The raw data file contains the data in logical record format (one line of data per respondent). The data are mapped on each line as follows: (a) columns 1-4 contain the ICPSR study number (6572); (b) columns 5-74 contain variables V01 to V70 with one column per variable; and (c) columns 75 to 81 contain the first weight variable (WT1). The WT1 variable should be included and any analysis should use the weighted data, as discussed above. Please note that if you create an SPSS or SAS file from the files provided, the data automatically will be weighted by WT1. If you wish for some reason to examine the unweighted data, it will be necessary to turn the weight factor off. The WT2 variable is not included in the raw dataset; it is computed by the SPSS or SAS commands.

Users who do not intend to use SPSS or SAS should still be able to access the data with minimal difficulty. Some statistical packages allow an SPSS file to be imported. Alternatively, one can work with the raw data file and use the information in the SPSS file definition statement file to create the appropriate control statements for the statistical package being used.

SETUPS: AMERICAN POLITICS

A SETUPS includes a monograph and a data set. SETUPS data are distributed by the Inter-University Consortium for Political and Social Research through an agreement with the American Political Science Association.
