## PROBLEM SET #5B: ANSWERS AND DISCUSSION

### General Comments

Evidently most students were able to produce SPSS frequency tables (and sometimes bar charts as well) without particular difficulty. Beyond thus, however, the product was usually less than satisfactory. Here are some problems that turned up more or less frequently.

1.  I may have done (or failed to do) two things (neither of which I can verify from my office) that may have caused problems. First, quite a few students evidently used 2009 rather than 2010 Student Survey data. It occurs to me that in Fall 2009 I may have posted Student Survey data in the Class Data => POLI 300 folder on the student PC network and then never removed it. (It is also possible that some students were copying information from last year's PS#5.A&D.) In any event, the current Student Survey data was distributed in class early in the semester and is also posted on the course website in electronic form in various formats. Second, with respect to Question 4, lno students included reported votes for "other" (than the Democratic or Republican) candidates. It also occurs to me that the Data7208.sav file on the student PC network may have code value "3" (as well as code value "9") set as missing data, thereby excluding "other" votes from tables. This can be changed by switching from "Data View" to "Variable View" and changing the Missing Values setting for V04, but it is understandable that students didn't do this.

2.  Often students provided little or no substantive conclusions based on the tables or charts they produced, or the discussion was not linked to specific attached tables or these tables were unclear because the variable and/or its values were not labelled.

3.  Questions #1 and #2 asked you to compare Student and ANES frequency distributions. While almost all students presented ANES frequencies in tables and/or charts (produced by SPSS), not all produced similar student frequency tables or charts to compare them with. Some students cited selected (and accurate) student percentages, thereby indicating that they had calculated such frequencies even though they were not comprehensively displayed in a table or chart. The following discussion shows what fully satisfactory answers to such questions look like. "Merged" frequency bar charts are an especially effective way of making comparisons such as these. For suggestions on how to produce such graphs using SPSS, see the end of the Answers & Discussion handout.

4.  In like manner, some students answered Question #4 by just asserting that ANES popular vote estimates closely matched "official" popular vote figures, without presenting such "official" figures for comparison. And many students who did present such "official" figures did not provide a source for this information.

5.  A general comparison of two frequency distributions should focus on the general *shape*, *location*, and *spread* of the distributions, not just selected frequencies (or bar heights).

6.  For Question #1, many all students compared student perceptions of the ideological position of the Republican Party in Spring 2009 with national (ANES) perceptions that extend from 1972 through 2008 (over which period the party arguably has become distinctly more

conservative). It would be more *valid* to compare students perceptions with ANES perceptions for 2008 (i.e., the most recent year available) only.

7.    Many students reached *invalid* conclusions (especially with respect to Question #3), because they focused on *absolute* rather than *relative* frequencies. For example, many students noted that the case counts (absolute frequencies) of Strong Democrats and Strong Republicans in most recent years are substantially smaller than in 1972 and 1976, and then took this as evidence that Party ID had weakened over the time period. But this decline in case counts is due entirely to smaller sample sizes (appearing as absolute marginal frequencies) in the later years. Calculation of relative frequencies (percentages that you can calculate by hand or have SPSS calculate and display by following the instructions is the SPSS handout) shows that Strong Democrats and Strong Republicans actually made up a slightly larger proportion of the samples in the later years.

8.    A surprising number of students simply used incorrect ANES variables, e.g., REP CAND IDEOLOGY (V36) or REP CAND THERMOMETER (V27) or even R's OWN IDEOLOGY (V34) rather than REP PARTY IDEOLOGY (V38) for Question #1.

## *Question1*

SETUPS/ANES variable V38 corresponds to Student Survey Q33. Here is the overall SPSS Frequency Tabulation for V38 (REPUBLICAN PARTY: IDEOLOGY).

### V38 REPUBLICAN PARTY: IDEOLOGY

|       |                  | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|------------------|-----------|---------|---------------|--------------------|
| Valid | liberal          | 1233      | 6.0     | 8.1           | 8.1                |
|       | slightly liberal | 1103      | 5.4     | 7.2           | 15.3               |
|       | moderate         | 2241      | 10.9    | 14.7          | 30.0               |
|       | slightly conserv | 3383      | 16.4    | 22.1          | 52.1               |
|       | conservative     | 7321      | 35.6    | 47.9          | 100.0              |
|       | Total            | 15281     | 74.2    | 100.0         |                    |
| Missing | NA             | 5302      | 25.8    |               |                    |
|       | Total            | 20583     | 100.0   |               |                    |

Bear in mind that this frequency distribution is based on *all ANES respondents from 1972 through 2008*. Perceptions of the ideological position of the Republican Party may have changed over this period of time (perhaps because the ideological position of the Republican party has actually changed). It is therefore more appropriate to compare student perceptions in 2009 with the most recent 2008 ANES data only. We can do this by using the Select Cases procedure, where we select cases such that "v01 = 2008." Even so we are comparing student and ANES respondents at different times and at different points in the electoral cycle.

### REPUBLICAN PARTY: IDEOLOGY (2008 ONLY)

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | liberal | 195 | 8.4 | 9.2 | 9.2 |
| | slightly liberal | 144 | 6.2 | 6.8 | 16.0 |
| | moderate | 304 | 13.1 | 14.3 | 30.3 |
| | slightly conserv | 386 | 16.6 | 18.2 | 48.5 |
| | conservative | 1094 | 47.1 | 51.5 | 100.0 |
| | Total | 2125 | 91.5 | 100.0 | |
| Missing | NA | 198 | 8.5 | | |
| | Total | 2323 | 100.0 | | |

Obviously absolute frequencies are vastly larger in the SETUPS/ANES data than in the Student Survey, so when we compare the frequency distributions for V38 and Q35, we should focus on *relative*, not *absolute*, frequencies (i.e., on percentages, not case counts).  Moreover, the ANES data (unlike the Student Survey data) includes a fair amount of missing data.  Therefore, we should compare *adjusted relative frequencies* (valid percents).  (Please examine the formatting in the following tables, which were created as part of this word processing document, and which illustrate typical "presentation grade" tables as might appear in a published article or book.)

### REPUBLICAN PARTY IDEOLOGY

| | *Students* | *ANES* | |
|---|---|---|---|
| *Perceived Ideology* | 2010 | 1972-2008 | 2008 only |
| *Liberal* | 8% | 8% | 9% |
| *Slightly liberal* | 0% | 7% | 7% |
| *Moderate* | 0% | 15% | 14% |
| *Slightly conservative* | 24% | 22% | 18% |
| *Conservative* | 68% | 48% | 52% |
| *Total* | 101% | 100% | 100% |
| | (*n* = 50) | (*n* = 15,281) | (*n* = 2125) |

Student respondents generally have a sharper perception of the Republican party as right-of-center than the general population does.  In particular, about 15% of the ANES respondents but only 8% of the students) claim to see the Republican party as at least slightly "liberal."  Most of these 15% of the ANES respondents probably have little or no understanding of the terms "liberal" and "conservative" and (despite a preceding filter question in the ANES questionnaire) are giving essentially random responses.  Conversely, two-thirds of the students but only about half of the national sample perceive the Republican to be clearly "conservative."

*Question 2*

SETUPS variable V36 is REPUBLICAN CANDIDATE IDEOLOGY.  If we select 2000 and/or 2004 cases only, the Republican candidate in question is George W. Bush.  (Most students did not "select cases" at all, so they were in effect comparing student perceptions of Bush in 2010 with national perceptions not only of Bush in 2000 and 2004 but also of Dole in 1996, Bush Sr. in 1992 and 1988, etc.)  We can select respondents from 2000 *and* 2004 by using a Select Cases expression such as "V01 > 1998 & V01 < 2006."

### V36 REPUBLICAN CANDIDATE: IDEOLOGY (2000 & 2004)

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | liberal | 264 | 9.6 | 10.9 | 10.9 |
| | slightly liberal | 146 | 5.3 | 6.0 | 16.9 |
| | moderate | 286 | 10.3 | 11.8 | 28.7 |
| | slightly conserv | 395 | 14.3 | 16.3 | 45.0 |
| | conservative | 1335 | 48.3 | 55.0 | 100.0 |
| | Total | 2426 | 87.7 | 100.0 | |
| Missing | NA | 340 | 12.3 | | |
| | Total | 2766 | 100.0 | | |

Of course, in 2000 George W. Bush had been a nationally recognized figure for only about a year and had not yet had much time to project any ideological image.   Subsequently, people had an opportunity to develop clearer perceptions.  It may therefore be more appropriate to compare student data with the 2004 ANES data only (though the differences between 2000 and 2004 are small and may reflect nothing but sampling error) .  Even doing this, we are comparing ANES responses pertaining to *an incumbent President running for re-election* with student responses pertaining to *an ex-President more than four years later.*

### V36 REPUBLICAN CANDIDATE: IDEOLOGY (2004)

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | liberal | 136 | 11.2 | 12.7 | 12.7 |
| | slightly liberal | 65 | 5.4 | 6.1 | 18.7 |
| | moderate | 113 | 9.3 | 10.5 | 29.2 |
| | slightly conserv | 139 | 11.5 | 13.0 | 42.2 |
| | conservative | 621 | 51.3 | 57.8 | 100.0 |
| | Total | 1075 | 88.7 | 100.0 | |
| Missing | NA | 137 | 11.3 | | |
| | Total | 1212 | 100.0 | | |

**PERCEIVED BUSH IDEOLOGY**

|  | *Students* | *ANES* | |
| --- | --- | --- | --- |
| *Perceived Ideology* | 2010 | 2000 and 2004 | 2004 only |
| *Liberal* | 4% | 11% | 13% |
| *Slightly liberal* | 4% | 6% | 6% |
| *Moderate* | 10% | 12% | 10% |
| *Slightly conservative* | 10% | 16% | 13% |
| *Conservative* | 72% | 55% | 58% |
| *Total* | 100%* | 100% | 100% |
|  | (*n* = 50) | (*n* = 2426) | (*n* = 1212) |

Again students have a much sharper perception of George W. Bush as conservative (in 2008) than ANES respondents have (even in 2004 only). A substantial minority of ANES respondents perceive (or, presumably, misperceive) Bush as some kind of liberal (probably for the same reasons as those noted in discussing Question 1), and even more students perceive Bush as "[strictly] conservative" than ANES respondents.

### *Question 3*

On the next page you will find an (edited) SPSS crosstabulation of V09 (PARTY IDENTIFICATION) by V01 (YEAR OF SURVEY), showing "column" [within year] percentages. (Note that it is *invalid* to compare *absolute frequencies* over the eight years because the ANES sample sizes vary quite a lot from year to year; for this reason, absolute frequencies have been edited out.) [SD = Strong Democrat, WD = Weak Democrat, DL = Independent Leaning Democratic, IND = Pure Independent, RL = Independent Leaning Republican, WR = Weak Republican, SR = Strong Republican, NP = Nonm-Partisan, L = Leaner, WP = Weak Partisan, SP = Strong Partisan]

It is hard to see much of a pattern in a table so large. Note that the proposition refers to *strength* of party identification without reference to the (Democratic or Republican) *direction* of that identification. What we can do is *recode* party identification by "folding" the range of values at its "Independent" midpoint — that is, by combining the SD and SR categories into a single "strong partisan" (SP) category, and so forth. (This specific example was discussed in Handout #4 and in class.) This can be done by either (i) using SPSS to *recode* V09 and running a new crosstabulation or (ii) simply *adding* the percentages in the rows in the above table together in the appropriate fashion to get the second table on the following page.

Just looking at the two extremes (1972 and 2008, we can see that Stronger Partisans increased and Non-Partisans decreased, but this take no account of the intermediate categories or the intermediate years. We can "boil down" the data still more to see more clearly what is (or is not) going on. The bottom line of the Strength of Party ID table above goes beyond the assignment in PS #5B, by displaying the *average (mean)* strength of party identification (to create a [not completely straight "line of averages," as on a scattergram) for each year, where Non-Partisan (Independent) = 1, Leaner = 2, Weak Partisan = 3, and Strong Partisan = 4. (We assume — perhaps without great justification — that STRENGTH OF PARTY IDENTIFICATION [recoded Party ID] forms an *interval* scale, so that we can average these numbers.)

**PARTY IDENTIFICATION BY YEAR OF SURVEY**

| PARTY ID | | YEAR OF SURVEY | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1972** | **1976** | **1980** | **1984** | **1988** | **1992** | **1996** | **2000** | **2004** | **2008** |
| | **SD** | 14.9% | 14.9% | 18.1% | 17.2% | 17.8% | 18.2% | 18.8% | 19.2% | 16.5% | 17.4% |
| | **WD** | 26.1% | 25.2% | 23.6% | 20.4% | 18.0% | 17.7% | 19.9% | 15.4% | 15.6% | 20.4% |
| | **DL** | 11.3% | 11.9% | 11.7% | 11.0% | 12.0% | 14.4% | 13.1% | 15.2% | 17.% | 16.2% |
| | **IND** | 13.3% | 14.7% | 13.2% | 11.1% | 10.8% | 11.7% | 8.0% | 11.8% | 9.7% | 8.5% |
| | **RL** | 10.6% | 9.8% | 10.5% | 12.6% | 13.5% | 12.5% | 11.1% | 13.5% | 11.7% | 10.9% |
| | **WR** | 13.3% | 14.5% | 14.3% | 15.0% | 14.1% | 14.2% | 15.8% | 12.1% | 12.5% | 13.7% |
| | **SR** | 10.5% | 9.0% | 8.7% | 12.6% | 14.0% | 11.2% | 13.4% | 12.9% | 16.5% | 14.1% |
| **Total** | | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| **n =** | | 2656 | 2826 | 1577 | 2198 | 1999 | 2448 | 2448 | 1508 | 1194 | 2123 |

**STRENGTH OF PARTY ID BY YEAR OF SURVEY**

| STRENGTH OF PARTY ID | | YEAR OF SURVEY | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1972** | **1976** | **1980** | **1984** | **1988** | **1992** | **1996** | **2000** | **2004** | **2008** |
| | **NP** | 13.3% | 14.7% | 13.2% | 11.1% | 10.8% | 11.7% | 8.0% | 11.8% | 9.7% | 8.5% |
| | **L** | 21.9% | 21.7% | 22.1% | 23.6% | 25.5% | 26.9% | 24.2% | 28.7% | 29.0% | 27.0% |
| | **WP** | 39.4% | 39.6% | 37.9% | 35.4% | 32.0% | 31.9% | 35.7% | 27.5% | 28.1% | 30.1% |
| | **SP** | 25.4% | 24.0% | 26.8% | 29.8% | 31.7% | 29.5% | 32.2% | 32.1% | 33.0% | 34.4% |
| **Total** | | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| **Count: n =** | | 2656 | 2825 | 1577 | 2198 | 1999 | 2448 | 1508 | 1525 | 1194 | 2123 |
| **Mean** | | **2.77** | **2.73** | **2.78** | **2.84** | **2.84** | **2.79** | **2.92** | **2.68** | **2.84** | **2.90** |

The picture displayed by the above table is more one of stability than change, in strength of party identification.  (The small fluctuations in the means can be attributed mostly to sampling error.)  Of course, there may be a question as to whether we have measured STRENGTH OF

PARTY IDENTIFICATION in an appropriate fashion.  (Despite the evidence above, there is some evidence that suggests that meaningful party identification indeed weakened from the mid-twentieth century until at least the mid-1990s — in particular,  that it was stripped of much "affective" [i.e., emotional, whether pro or con] content; see Martin Wattenberg, *The Decline of American Political Parties: 1952-1996*.)

The best way to depict this change or stability over time graphically would be by means of a *stacked bar* (or *line*) *chart*.

*Question 4*

We crosstabulate V04 (PRESIDENTIAL VOTE) with V01 (YEAR OF SURVEY) and examine column [within year] percentages, after excluding missing data (nonvoters).  Here is the crosstabulation, using the ANES 1972-2004 data set.  Many students did not include the "other" category (see Note 1 on the first page of this A&D), which meant they made *invalid* comparisons between the ANES popular vote percentages and "official" percentages that include support for third candidates, notably in 1980, 1992 and 1996, when John Anderson and Ross Perot captured significant percentages of the popular vote.  (I may have inadvertently submitted to OIT a SETUPS file in which "3" as well as "9" was designated missing data.)

## PRESIDENTIAL VOTE BY YEAR OF SURVEY

| PRES VOTE | YEAR OF SURVEY | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1972 | 1976 | 1980 | 1984 | 1988 | 1992 | 1996 | 2000 | 2004 | 2008 |
| **Dem** | 35.3% | 50.2% | 39.4% | 41.4% | 46.6% | 47.7% | 52.7% | 50.1% | 49.1% | 53.7% |
| **Rep** | 63.6% | 47.9% | 50.8% | 57.7% | 52.3% | 33.8% | 39.7% | 45.9% | 49.3% | 44.3% |
| **Other** | 1.1% | 1.9% | 9.8% | 0.9% | 1.2% | 18.4% | 7.6% | 4.0% | 1.6% | 2.0% |
| **Total** | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| **n =** | 1605 | 1663 | 972 | 1389 | 1209 | 1661 | 1121 | 1098 | 799 | 1570 |

The next step is to track down the "official" popular vote percentages for the Democratic and Republican Presidential candidates in each of the eight elections.  For a variety of reasons, there may be slight discrepancies among different reference sources but they do not amount to more than a fraction of a percent.  Since the ANES table above also shows vote percentages for other ("minor") Presidential candidates, you should take care to use a reference source that likewise puts the popular vote percentages on a total vote, rather than two-party vote, basis.  The following "official" popular

vote figures for 1972-2004 come from the *Congressional Quarterly's Guide to U.S. Elections* (available in the AOK Library reference room) and also at *David Leip's Atlas of U.S. Elections*, a very useful website (at **http://uselectionatlas.org/**).

**Popular Vote for President**

|  | *Official Report* | | *ANES Samples* | |
|---|---|---|---|---|
|  | *Dem. %* | *Rep. %* | *Dem. %* | *Rep. %* |
| 1972 | 37.5 | 60.7 | 35.3 | 63.6 |
| 1976 | 50.1 | 48.0 | 50.2 | 47.9 |
| 1980 | 41.0 | 50.8 | 39.4 | 50.8 |
| 1984 | 40.6 | 58.8 | 41.4 | 57.7 |
| 1988 | 45.6 | 53.4 | 46.6 | 52.3 |
| 1992 | 43.3 | 37.7 | 47.7 | 33.9 |
| 1996 | 49.2 | 40.7 | 52.7 | 39.7 |
| 2000 | 48.4 | 47.8 | 50.1 | 45.9 |
| 2004 | 48.3 | 50.7 | 49.1 | 49.3 |
| 2008 | 52.9 | 45.6 | 53.7 | 44.3 |

If you could not figure out how to get ANES percentages based on the total Presidential vote (including for "other" candidates), you should have adjusted the "official figures" to likewise exclude "other" candidates.

In general, the figures match very well. It is evident that the ANES estimates are subject to (a) *sampling error* of a few percentage points (as we know is true of all random samples) and, possibly, (b) perhaps a slight *bias* in favor of the winner due (not to sampling error but) to misreporting by a small number of respondents subject to "bandwagon effects" (i.e., a tendency for some respondents to claim that they voted for the winner when in fact they did not) in post-election interviews (especially prominent in 1972, 1992, and 1996). However, 2000 was a conspicuous exception to this rule, probably due to the fact that (most if not all) the ANES post-election interviews took place *before* the election dispute in Florida was resolved, so it remained uncertain who the winner was.

### *Merged Bar Charts*

Merged bar charts for Questions #1 and #2 are shown on the next pages.  Some students presented SPSS bar charts for the ANES data in Questions #1 and #2, and some also presented separate bar charts for the student data, evidently produced by Excel or some similar spreadsheet program.  Here is how to construct the SPSS merged bar charts shown on the following page.

After getting the relative frequencies for both data sets, open a blank SPSS data entry screen by clicking on File => New => Data.  Then enter the frequency data as shown below. Note that *code values* of the IDEOLOGY variables becomes  *cases* (rows) and the *relative frequencies* become *variables* (columns). (I continue to round to whole percentage points.)

| case # | ssrep | anesrep | ssgwb | anesgwb |
|--------|-------|---------|-------|---------|
| 1 | 8 | 9 | 4 | 13 |
| 2 | 0 | 7 | 4 | 6 |
| 3 | 0 | 14 | 10 | 10 |
| 4 | 24 | 18 | 10 | 13 |
| 5 | 68 | 52 | 72 | 58 |

Using this data file, one can create merged bar charts by clicking on Graphs => Legacy Dialogs => Bar => Clustered where Data in Chart are Values of individual cases.  Then select ssrep and anesrep for Bars Represent for the first chart and ssgwb and anesgwb for the second. You can leave the radio button on Case Number for Category Label.  If you double click on an SPSS chart in the output screen, you will get an editing screen with different menus that can be use to modifying the labels and formatting and/or you can export the SPSS chart (File => Export => Chart Only: Windows Bitmap) to Windows Paint or some fancier drawing program to embellish it.

``