

FREQUENCY DISTRIBUTIONS: ANSWERS AND DISCUSSION

- Any frequency table and/or bar graph should have a heading or title that clearly identifies the *population* whose data is displayed (e.g., students vs. ANES) and the *variable* of interest. Likewise the rows of the table, or the bars in the graph, that represent the possible *values* of the variable should be clearly labeled (not with code values but with actual words for qualitative variables or actual numbers for quantitative variables). Otherwise, your reader cannot understand the information you are trying to convey.

Copies of the requested SETUPS frequency tables as produced by SPSS shown here. Your (probably hand-drawn) tables should be set up in more or less the same way, and the percentages you entered (presumably using a calculator) should match those displayed by SPSS. (However, rounding to the nearest whole percentage point is fully adequate.)

V34 R'S OWN IDEOLOGY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	liberal	291	12.5	16.7	16.7
	slightly liberal	205	8.8	11.8	28.5
	moderate	506	21.8	29.1	57.6
	slightly conserv	278	12.0	16.0	73.5
	conservative	461	19.8	26.5	100.0
	Total	1740	74.9	100.0	
Missing	NA	583	25.1		
	Total	2323	100.0		

V42 HEALTH PLAN SCALE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	government plan	363	15.6	34.3	34.3
	2	140	6.0	13.2	47.5
	3	200	8.6	18.9	66.4
	4	136	5.9	12.9	79.3
	private plan	219	9.4	20.7	100.0
	Total	1058	45.6	100.0	
Missing	NA	1265	54.4		
	Total	2323	100.0		

V45 ABORTION OPINION

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never permit	154	6.6	14.7	14.7
	for rape, etc.	285	12.3	27.3	42.0
	if need established	187	8.1	17.9	59.9
	always permit	419	18.0	40.1	100.0
	Total	1045	45.0	100.0	
Missing	NA	1278	55.0		
	Total	2323	100.0		

Review Points. The number in the *Frequency* (also called *Absolute Frequencies*) column is the actual count of cases that have the value corresponding to that row of the table, e.g., 291 respondents gave the “liberal” responses to the IDEOLOGY question, 205 gave the “slightly liberal response,” etc. Notice that a total of 1740 respondents gave a “valid” response to this question; an additional 583 cases are *missing* on this variable (they did not answer, or were not asked, the IDEOLOGY question), for an overall total of 2323 cases. The number in the *Percent* (or *Relative Frequency*) column is the absolute frequency as a percent of the overall number of cases, e.g., $291/2323 = 0.125 = 12.5\%$. The number in the *Valid Percent* (or *Adjusted Relative Frequency*) column is the absolute frequency as a percent of the number of *valid* (or *non-missing*) cases, e.g., $291/1740 = 0.167 = 16.7\%$. The number in the *Cumulative Percent* column is the cumulative sum of the valid percents moving downward (e.g., from “liberal” to “conservative”), e.g., in the “slightly liberal” row, $28.5\% = 16.7\% + 11.8\%$, and the cumulative percent in the last row is always 100%. (One can also cumulate in the other direction.) Such cumulative frequencies make sense in these frequency tables because all three variable are *ordinal* in nature.

Note. There was an error in the data you were given. ANES Missing Data on Health Plan Scale is actually 1265 (not 1058 you were given, which is rather the total number of valid cases), so your (Unadjusted) Relative Frequencies will not match those shown here.

2. The corresponding variables in the Fall and Spring 2009 Student Survey are Q24, Q14, and Q13 respectively. Copies of the requested Student Survey frequency tables as produced by SPSS are shown here. Again, your (probably hand-constructed) tables should be set up in more or less the same way, and display similar numbers.

V24 R's OWN IDEOLOGY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Liberal	13	26.0	26.0	26.0
	Slightly Lib	12	24.0	24.0	50.0
	Moderate	10	20.0	20.0	70.0
	Slightly Cons	9	18.0	18.0	88.0
	Conservative	6	12.0	12.0	100.0
	Total	50	100.0	100.0	

V14 HEALTH INSURANCE SCALE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Government	10	20.0	20.0	20.0
	Mostly Govt	12	24.0	24.0	44.0
	Mixed	11	22.0	22.0	66.0
	Mostly Private	12	24.0	24.0	90.0
	Private	5	10.0	10.0	100.0
	Total	50	100.0	100.0	

V13 ABORTION OPINION

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	For Rape, etc.	14	28.0	28.0	28.0
	If Need Established	6	12.0	12.0	40.0
	Always Permit	30	60.0	60.0	100.0
	Total	50	100.0	100.0	

Note 1. Making tally marks is part of the clerical process of *constructing* a table, but tally marks (and value codes) really should not be displayed in a finished product, e.g., in a table appearing in a paper. There is no problem in displaying tally marks (and value codes) in POLI 300 problem sets, but you should also count up the tallies and display these counts as *absolute frequencies*.

Note 2. By default and I think unfortunately, SPSS does not display rows that have no cases, e.g., “Never Permit” in student ABORTION OPINION.

Note 3. Missing data should (almost) always be excluded from bar (and other) charts. It is especially important to do this when making comparisons between POLI 300 student vs. ANES respondents, because the national data always includes quite a bit of missing cases while the student data has few or no missing cases (in fact, none at all in this data, so *valid percent* = *percent* in all three tables).

Note 4. Several students seem to regard the intermediate categories on the Health Plan Scale as missing data, presumably because the SETUPS Codebook gives verbal labels only for the two ends of the scale. But the intermediate values are valid data and should be included in tables and charts.

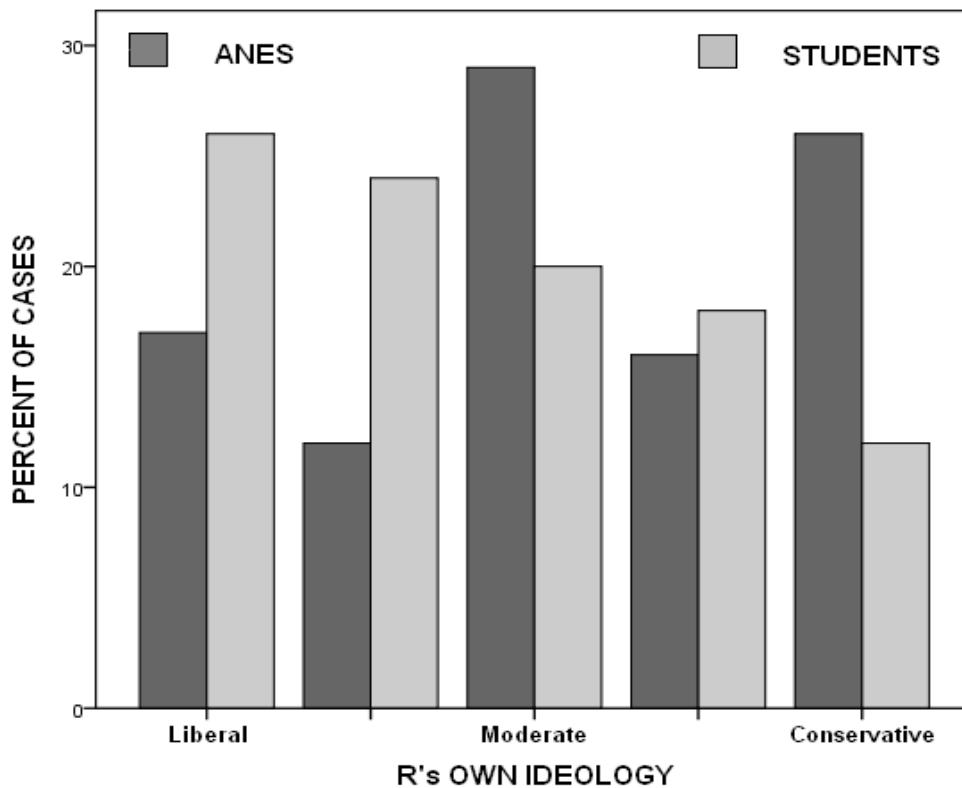
Note 5. A common problem throughout was that quite a few students turned in tables and (especially) charts that were not *labeled*, with the result that no reader (other than me) could figure out what the table or chart is saying. You need to specify the *data* being presented (here either *students* or *ANES*), the *variable* whose distribution is being displayed (e.g., “Respondent’s Own Ideology,” not “V34”), and the values of the variable (e.g., “Liberal,” not the code value “1”).

Preliminary Points Concerning Student/ANES Comparisons

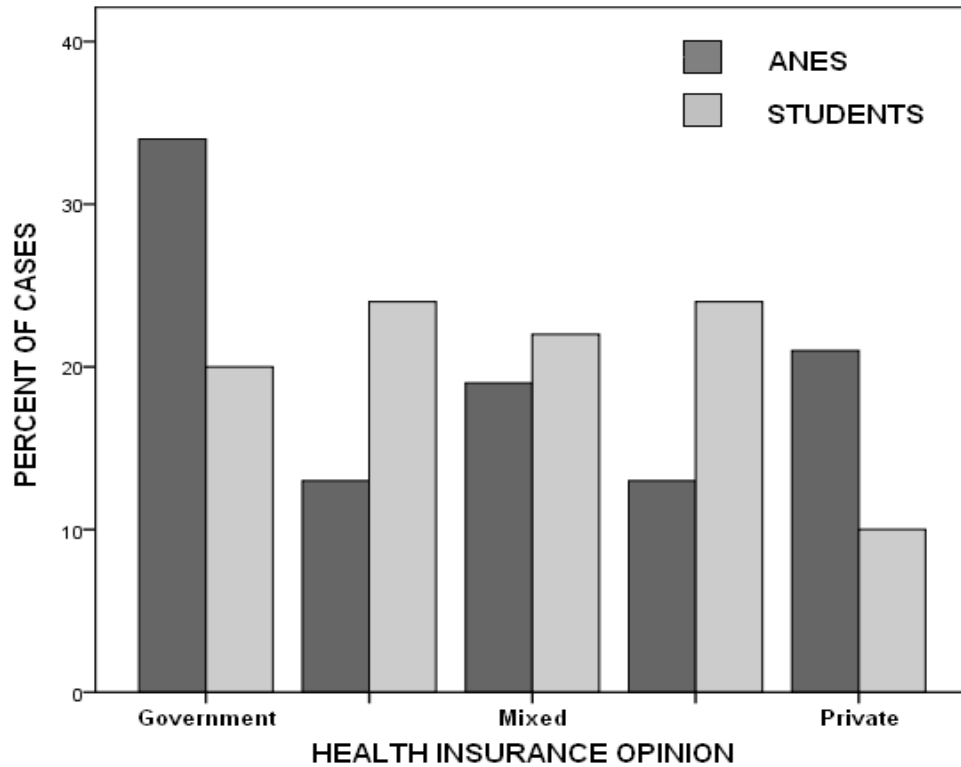
- (a) In comparing the SETUPS/NES and Student Survey frequency distributions of the parallel variables, you should of course compare their respective *adjusted relative frequencies* (or *valid percents*). (Obviously absolute frequencies are always far larger in the SETUPS data; also, as noted above, the SETUPS data has a lot of missing data, while the Student Survey has none.)
- (b) Many students, having gone to considerable trouble to produce the comparable student and ANES frequency tables and bar graphs, then made no effort actually to compare them and note ways in

which the two sets of distributions may be similar or different. The whole point of constructing tables and charts (and doing data analysis in general) is *to reach substantive conclusions that can be stated in words*. And many students who did try to make comparative statements were pretty vague. Another common problem was to focus only on which student and ANES bars were about the same height and which were of clearly different heights. This is an obvious first step but the comparison should be *general*, focusing on *the student and ANES frequency distributions as a whole*, not merely on the relative frequencies of particular values.

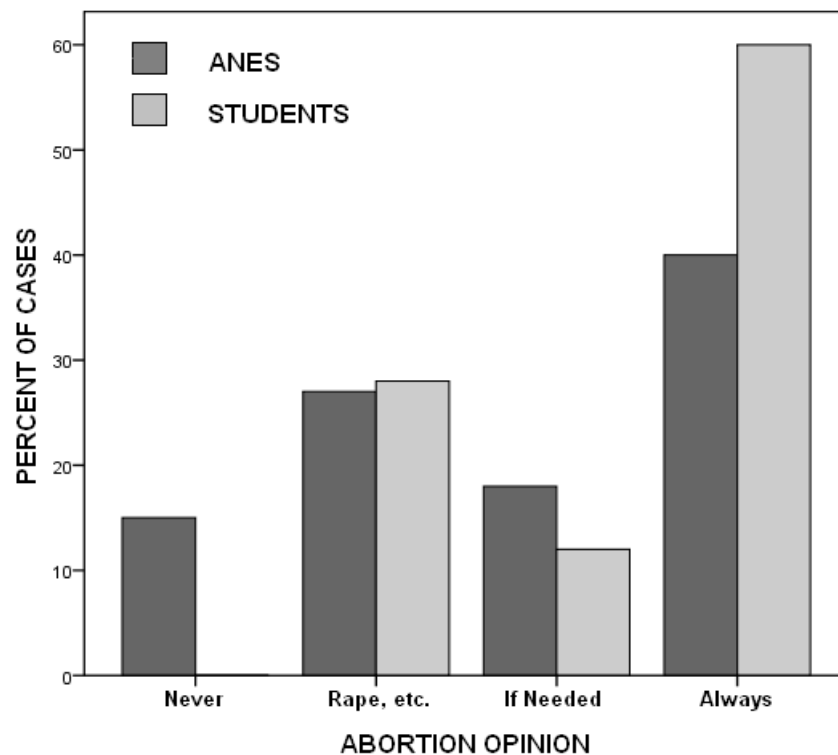
The following points can be observed when ANES and student relative frequencies are compared. *Merged bar graphs* for the student and ANES data facilitate making such comparisons and are shown below. (See end of Answers & Discussion on how these were constructed.)



- A. ***Ideology.*** The student sample leans distinctly to the liberal side, while the ANES sample leans more to the conservative side, with a large proportion of moderates.



- B. **Health Plan.** Despite their (relatively) “conservative” tendencies with respect to (self-reported) ideology, the ANES sample is actually considerably more favorable toward a strictly “government” health plan than the students are. But the ANES is also considerably more favorable toward a strictly “private” plan than the students are. In general, students tend to be more “middle of the road” on this issue (though leaning to the “left”), while the ANES sample is more *polarized* between the government vs. private extremes.



- C. **Abortion.** Both the students and ANES respondents lean quite heavily in the “pro-choice” direction, but this tendency is especially pronounced among students. In particular, no students are strictly “pro-life” (“Never”) and many more are strictly “pro-choice.”

Methodological Note. These merged/clustered bar graphs were created by SPSS applied to a small aggregate data spreadsheet with 5 cases and 7 variables. The 5 “cases” are the five possible values of each variable; the first variable is the code value (1 through 5), and the remaining variables are the corresponding *adjusted relative frequencies (valid percents)* for IDEOLOGY, HEALTH PLAN, and ABORTION for the students (V2 = Q24, V4 = Q14, and V6 = Q13) and the ANES sample (V3 = V34, V5 = V42, and V7 = V45) and respectively. These numbers can be typed directly into the SPSS data editor screen (that is the first thing you see when you load SPSS) to produce a small data file that looks like this:

<u>V1</u>	<u>V2</u>	<u>V3</u>	<u>V4</u>	<u>V5</u>	<u>V6</u>	<u>V7</u>
1	26	17	20	34	0	15
2	24	12	24	13	14	27
3	20	29	22	19	6	18
4	18	16	24	13	30	40
5	12	26	10	21	.	.

I then requested *Clustered Bar Charts* where *Data in Chart are Values of Individual Cases*. I also used *Select Cases* for ABORTION such that “V1 < 5,” so as to suppress a fifth pair of bars with zero height. I then *exported* the SPSS charts to Windows Paint for further editing and annotation.) Some of you submitted computer bar charts evidently created in much the same way, perhaps using Excel or a similar spreadsheet program.