

FREQUENCY DISTRIBUTIONS: ANSWERS AND DISCUSSION

1. 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 simply 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	133	11.0	14.7	14.7
	slightly liberal	94	7.8	10.4	25.1
	moderate	304	25.1	33.5	58.5
	slightly conserv	144	11.8	15.8	74.4
	conservative	232	19.2	25.6	100.0
	Total	907	74.8	100.0	
Missing	NA	305	25.2		
Total		1212	100.0		

V42 HEALTH PLAN SCALE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	government plan	372	30.7	33.6	33.6
	2	145	12.0	13.1	46.7
	3	209	17.2	18.9	65.6
	4	162	13.4	14.7	80.3
	private plan	218	18.0	19.7	100.0
	Total	1106	91.3	100.0	
Missing	NA	106	8.7		
Total		1212	100.0		

V45 ABORTION OPINION

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	never permit	149	12.3	14.2	14.2
	for rape,etc.	335	27.6	32.0	46.2
	if need established	186	15.4	17.8	64.1
	always permit	376	31.0	35.9	100.0
	Total	1046	86.3	100.0	
Missing	NA	166	13.7		
Total		1212	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., 133 respondents gave the “liberal” responses to the IDEOLOGY question, 94 gave the “slightly liberal response,” etc. Notice that a total 907 respondents (907 “valid cases”) gave a “valid” response to this question; an additional 305 cases are *missing* on this variable (they did not answer, or were not asked, the IDEOLOGY question), for an overall total of 1212 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., $133/1212 = 0.110 = 11.0\%$. 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 (907 in the IDEOLOGY table), e.g., $133/907 = 0.147 = 14.7\%$. The number in the *Cumulative Percent* column is the cumulative sum of the valid percents moving downward (from “liberal” to “conservative”), e.g., in the “slightly liberal” row, $25.1\% = 14.7\% + 10.4\%$, and the cumulative percent in the last row is always 100%. (One can also cumulate in the other direction.)

2. Copies of the requested Student Survey frequency tables as produced by SPSS are shown here. Again, your hand-constructed tables should be set up in more or less the same way, and display similar numbers. (For some reason, some students used V01 (PARTY ID), rather than V24 (R’s OWN IDEOLOGY.)

R’s OWN IDEOLOGY

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Liberal	14	29.2	29.2	29.2
	Slightly Lib.	8	16.7	16.7	45.8
	Moderate	14	29.2	29.2	75.0
	Slightly Cons.	3	6.3	6.3	81.3
	Conservative	8	16.7	16.7	97.9
	NA	1	2.1	2.1	100.0
	Total	48	100.0	100.0	

HEALTH PLAN SCALE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Government	15	31.3	31.3	31.3
	Mostly govt.	5	10.4	10.4	41.7
	Mixed	15	31.3	31.3	72.9
	Mostly private	8	16.7	16.7	89.6
	Private	5	10.4	10.4	100.0
	Total	48	100.0	100.0	

ABORTION OPINION

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never permit	4	8.3	8.5	8.5
	For rape, etc.	11	22.9	23.4	31.9
	If need established	7	14.6	14.9	46.8
	Always permit	25	52.1	53.2	100.0
	Total	47	97.9	100.0	
Missing	NA	1	2.1		
Total		48	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 exercises, but you should also count up the tallies and display these counts as *absolute frequencies*.

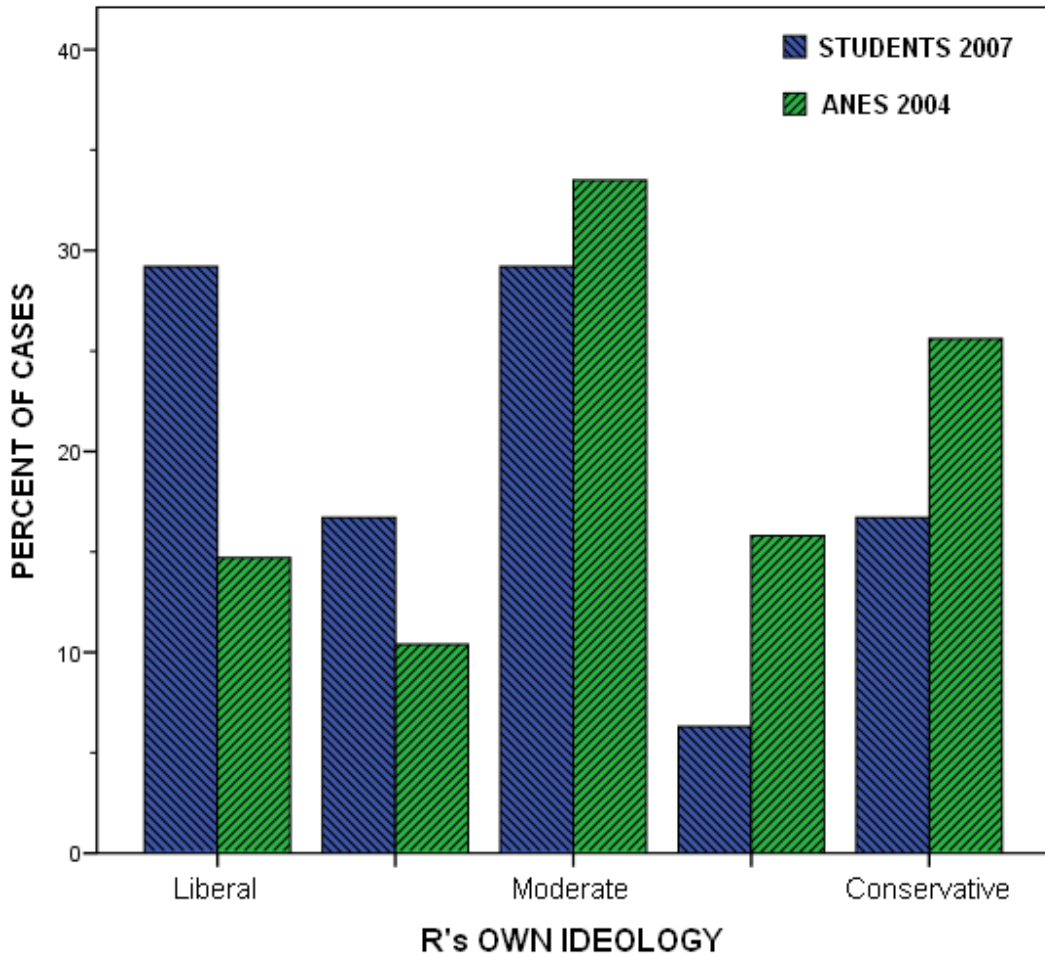
Note 2. Missing data should (almost) always be excluded from bar (and other) charts. It is especially important to do this when making POLI 300 student vs. ANES comparison, because the national data always includes quite a bit of missing cases while the student data has few or no missing cases.

Note 3. 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.

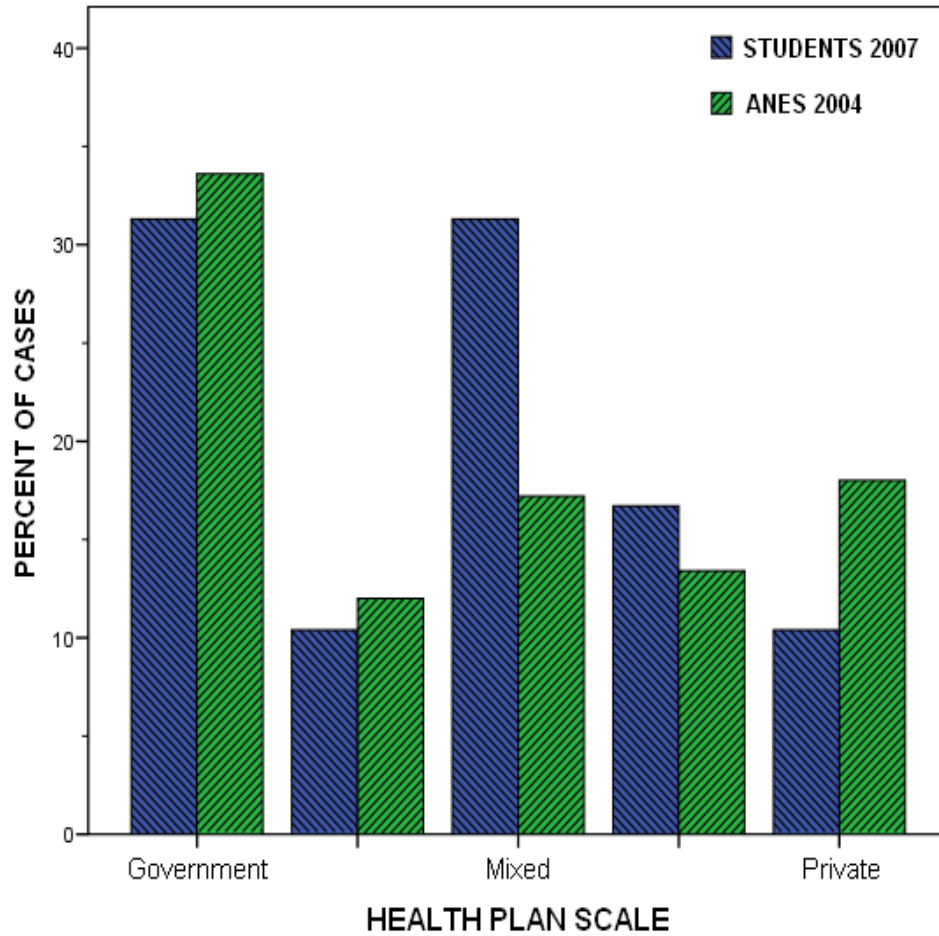
Preliminary Points Concerning Student/NES Comparisons

- (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 almost none.)
- (b) Many students, having gone to considerable trouble to produce the comparable student and NES 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 NES 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 NES frequency distributions as a whole*, not merely on the relative frequencies of particular values.

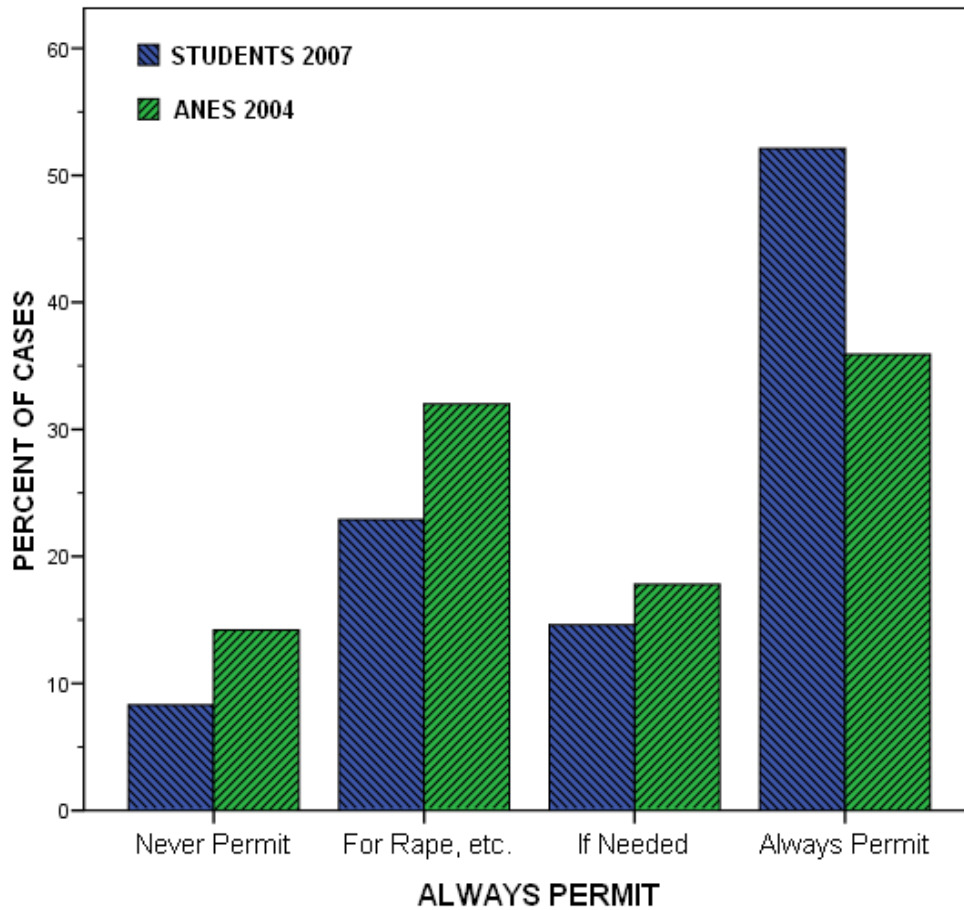
The following points can be observed when NES and student relative frequencies are compared. *Merged bar graphs* for the student and NES data facilitate making such comparisons and are shown below. (See end of Answers & Discussion on how these were constructed.)



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



B. **Health Plan.** Despite their (relatively) conservative tendencies with respect to ideology, the ANES sample is just as favorable toward a strictly “government” health plan as the students are. At the same time, the ANES is also more favorable toward a strictly “private” plan than the students. In general, students tend to be more “middle of the road” on this issue (though leaning heavily 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 in the “pro-choice” direction, but this tendency is much more pronounced among students than ANES respondents. **Note.** Some students speculated that there might be a “gender gap” on ABORTION OPINION. We cannot test this hypothesis in the student sample data (since no demographic data was collected) but can do so in the ANES data. See Slide 26 in *Cumulative Student Survey (plus ANES) Charts* [PPT] on the POLI 300 web page.

Methodological Note. The attached merged bar graphs were created by SPSS applied to a small *aggregate data* spreadsheet with 5 *cases* and 7 *variables*. The 5 cases are the possible variable values; the first variable is the code value (1 through 5), and the remaining variables are the corresponding *relative frequencies* for IDEOLOGY, HEALTH PLAN, and ABORTION for students (V2, V4, and V6) and the NES (V3, V5, and V7) 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	29.2	14.7	31.3	33.6	8.5	14.2
2	16.7	10.4	10.4	12.0	23.4	32.0
3	29.2	33.5	31.3	17.2	14.9	17.8
4	6.3	15.8	16.7	13.4	53.2	35.9
5	16.7	25.6	10.4	18.0	.	.

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 < 4,” so as to eliminate a fifth pair of bars both with zero height. I then *exported* the SPSS charts to Paint for further editing and annotation.) Some of you submitted computer bar charts evidently created in much the same way, probably using Excel or a similar spreadsheet program.