

PROBLEM SET #10: ANSWERS AND DISCUSSION***General Comments***

1. Remember that the convention is that the independent variable is made the column variable and the dependent variable is made the row variable. The “coloring” hypothesis implies that PARTY ID ==> PRES APPROVAL so, following the convention, PARTY ID should be column variable and PRES APPROVAL should be row variable. Your own hypothesis should likewise identify an independent and dependent variable and follow the same convention.
2. Tally marks and/or case IDs are part of the process of constructing a crosstabulation. The finished table should show case counts (absolute frequencies), and not (only) tallies or IDs. Tallies and/or IDs would certainly be removed from a “presentation-grade” version of a table (see p. 3), as might be included in a term paper, article, or book.
3. Some students “recoded” variables into *less precise* measures (e.g., by combining “Strongly Approve with “Approve”), and some simply dropped intermediate values (e.g., Independent [on PARTY ID], Moderate [on IDEOLOGY], etc.). It is best to keep the more precise measures in an initial crosstabulation, though values such as NA, DK, etc., and perhaps others are likely to be excluded from presentation grade tables (again see p. 3.)
4. Quite a few students constructed tables but then did not use them to reach any conclusions about the substantive proposition or hypothesis they were.

Questions

1. You should produce a table set up (preferably making the *independent* variable PARTY ID the *column* variable and the *dependent* variable PRES APPROVAL the *row* variable) more or less in the manner shown on the top of the next page, with case counts (absolute frequencies) shown in each cell *and with both variables and values clearly labeled*.

OVER =>

2. Different people did different tables, so no general answer can be provided (and I did not verify your table [using SPSS] unless something appeared to be clearly wrong). But in any case *your table again should be appropriately set up and labeled*. One quite common problem is that students crosstabulated variables at least one of which has in fact have *very little variability* or *dispersion* in this data (e.g., follow public affairs, voted in 2008; perceive the George W. Bush or Republican Party as conservative, etc.). If there is little or no variability in a variable, it can have little or no association with any other variable.

TABLE 1: PRESIDENTIAL APPROVAL BY PARTY IDENTIFICATION
(Fall 2010 Student Survey)
PARTY IDENTIFICATION (Q1)

PRESIDENTIAL APPROVAL (Q12)		Dem 1	Ind 2	Rep 3	Othr 4	DK 5	NA 9	Row Total
Strongly Approve	1	3	1	0	0	0	0	4
Approve	2	17	2	2	0	1	0	22
Disapprove	3	1	5	7	1	1	0	15
Strongly Disapprove	4	0	0	3	0	0	0	3
No Opinion	5	0	3	3	0	0	0	6
NA	9	0	0	0	0	0	0	0
Column Total		21	11	15	1	2	0	50

Note 1. This table is set up as a *worksheet*, not a *finished product*. The shaded rows and columns are included to allow for every possible combination of values on the two variables that may appear in the data. They would most likely be removed from a finished table appearing in a paper, article, or book. (Value codes would also be removed.)

Note 2. As a check, the *marginal frequencies* (row and column totals) should match the (univariate) absolute frequencies for Q1 and Q12. You should understand that if the two sets of frequencies fail to match, something is wrong with your table; however, even if the two frequencies do match, your table may still be wrong (because tables with different counts in the [interior] cells can have the same marginal frequencies; see Handout #10, Tables 1A-F).

Note 3. When you construct a crosstabulation (with a small number of cases) by hand like this, it can be helpful to do the following. Instead of just putting a tally marks in each cell for each case that belongs in that cell, actually write the ID number of that case in the cell. For example, when you are done, you will know not only that you have put *one case* in the row 2, column 4 [Disapprove, Other/Minor] cell of the table but also that it is the *case with the ID number 22* that belongs in this cell. This facilitates checking for clerical errors (for example, if you find the kind of discrepancy mentioned in Note 2). *But when you are done, you should enter the case counts (absolute frequencies) into each cell.*

Note 4. There is no reason to throw away information by using less precise measures. For example, use all four categories of PRESIDENTIAL APPROVAL, i.e., strong vs. weak (dis)approval. Likewise, keep the Independent category of PARTY ID (see if their views are “between” those of partisans). You can always combine or delete categories later.

Note 5. The relationship between the two variables can be seen more clearly if we examine *column percentages* (adjusted relative frequencies). (You were *not* asked to do this in PS #10, but we take

up table percentaging in Topic #12 .) In doing this, we should omit the [shaded] *missing data* cells (code 9) for both variables (i.e., we should calculate *adjusted relative frequencies/valid percentages*) and probably also omit the [also shaded] “residual categories” (codes 4 and 5) of PARTY IDENTIFICATION. A more difficult question (for which there is no clearly correct answer) is whether code 5 of V05 (“No opinion”) should be regarded as *missing data* to be excluded from the column percent table or whether it should be regarded as a *legitimate “neutral” category* of the *ordinal* variable PRESIDENTIAL APPROVAL that lies *between* “Approve” and “Disapprove.” Here I choose the latter option.

**TABLE 3: PRESIDENTIAL APPROVAL BY PARTY IDENTIFICATION:
FALL 2010**

<i>PRESIDENTIAL APPROVAL</i>	<i>PARTY IDENTIFICATION</i>		
	<i>Dem</i>	<i>Ind</i>	<i>Rep</i>
<i>Strongly Approve</i>	14%	9%	0%
<i>Approve</i>	81%	18%	13%
<i>No Opinion</i>	0%	27%	20%
<i>Disapprove</i>	5%	45%	47%
<i>Strongly Disapprove.</i>	0%	0%	20%
	100%	99%*	100%
	(n=21)	(n=11)	(n=15)

Source: POLI 300 Student Survey, Fall 2010

* Rounding error

Such a column percent table facilitates comparison of the distribution of dependent variable (PRESIDENTIAL APPROVAL) values across different categories of the independent variable (PARTY ID). Certainly the table strongly supports the “coloring” hypothesis, in so far such a small sample can support anything. For such a column percent table, it is especially important to report the number of cases in each column, because all the numbers are so small.

For comparison, here are the crosstabulations from last year and also the Fall 2008 and Fall 2007 Student Surveys pertaining to President Bush.

**TABLE 3: PRESIDENTIAL APPROVAL BY PARTY IDENTIFICATION:
FALL AND SPRING 2009**

<i>PRESIDENTIAL APPROVAL</i>	<i>PARTY IDENTIFICATION</i>		
	<i>Dem</i>	<i>Ind</i>	<i>Rep</i>
<i>Strongly Approve</i>	20%	15%	0%
<i>Approve</i>	66%	50%	20%
<i>No Opinion</i>	14%	25%	60%
<i>Disapprove</i>	0%	5%	0%
<i>Strongly Disapprove.</i>	0%	5%	20%
	100%	100%	100%
	(n=35)	(n=20)	(n=5)

Source: POLI 300 Student Survey, Fall 2010

The 2009 table certainly does not contradict the “coloring” hypothesis, but it does not provide strong support either, for two reasons:

- (a) the independent variable is substantially skewed in the “Democratic” direction; and
- (b) the dependent variable is highly skewed in the “approve” direction.

We can’t learn much about the association between variables if they don’t actually vary much in our data. But in so far as we can see any association, it is consistent with the “coloring” hypothesis: Democrats are (even) more approving than Independents and Republicans, and Independents are (on average) more approving than the few Republicans. But if there were more Republican cases, we might find them most of the approve of President Obama also, so in fact there would be little or no association between variables.

For comparison, here are the similar tables for President Bush’s last two years in office.

**TABLE 4: PRESIDENTIAL APPROVAL BY PARTY IDENTIFICATION:
FALL 2008**

<i>PRESIDENTIAL APPROVAL</i>	<i>PARTY IDENTIFICATION</i>		
	<i><u>Dem</u></i>	<i><u>Ind</u></i>	<i><u>Rep</u></i>
<i>Strongly Approve</i>	0%	0%	8%
<i>Approve</i>	0%	11%	42%
<i>No Opinion</i>	4%	11%	33%
<i>Disapprove</i>	28%	33%	8%
<i>Strongly Disapprove.</i>	68%	44%	8%
	100%	99%*	99%*
	(n=25)	(n=9)	(n=12)

Source: POLI 300 Student Survey, Fall 2008

* Rounding error

**TABLE 5: PRESIDENTIAL APPROVAL BY PARTY IDENTIFICATION:
FALL 2007**

<i>PRESIDENTIAL APPROVAL</i>	<i>PARTY IDENTIFICATION</i>		
	<i><u>Dem</u></i>	<i><u>Ind</u></i>	<i><u>Rep</u></i>
<i>Strongly Approve</i>	0%	8%	0%
<i>Approve</i>	0%	0%	70%
<i>Disapprove</i>	38%	23%	20%
<i>Strongly Disapprove.</i>	62%	69%	10%
	100%	99%*	100%
	(n=24)	(n=13)	(n=10)

Source: POLI 300 Student Survey, Fall 2007

* Rounding error