

**DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Research Methods Posc 302**

ANALYSIS OF SURVEY DATA

I. TODAY'S SESSION:

- A. Second steps in data analysis and interpretation
 - 1. Examples and explanation

II. RECAP:

- A. We want to investigate the relationship between
 - 1. Gender and attitudes toward defense spending and health care and
 - 2. Positions on issues and candidate choice.
- B. The purpose is to provide some background on public opinion for campaign strategists.
- C. Method:
 - 1. Secondary analysis of the 1996 American National Election Study.
 - 2. Cross-tabulations
 - i. We need information of the following sort.

	Men	Women
For	?	?
Not sure	?	?
Against	?	?

Figure 1: Plan Ahead - What Will the Analysis Look Like?

III. GENERATING CROSS-TABULATION TABLES:

- A. Getting the tables is relatively easy if you follow a few simple steps



1. Tip: as mentioned in Class 8 open two browser windows so that you have a copy of the codebook in one and the “crosstabs” dialogue in the other.
 - i. You can then copy and paste variable numbers.
 2. Tip: be careful to keep track of dependent and independent variables.
 - i. Treat the **independent** variable as the column (horizontal) variable.
- B. Go to the main page of the National Election study and check “Frequencies or cross tabulation” and click start.
1. You’ll see a dialog box like the one shown in the next figure.

SDA Tables Program
(Selected Study: 1996 American National Election Study)
Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify

Row:

OPTIONAL Variable names to specify

Column:

Control:

Selection Filter(s): Example: age(18-50) gender(1)

Weight:

Percentaging: Column Row Total

Other options

[Statistics](#) [Suppress table](#) [Question Text](#)

[Color coding](#) [Show T-statistic](#)

Figure 2: Cross-tabulation Dialog Box

2. It’s simple to make a cross-tabulation or contingency table if you take your time and make sure variable names are correctly typed and enter in the proper place.
 - i. I prefer that the dependent variable be the row variable and the independent variable is the column.
 - ii. It’s also going to be important to obtain column percentages in order to simplify the interpretation of the data.
 - iii. To avoid mistakes I sometimes use cut and past **variable numbers only** from Notepad to the appropriate dialog box entry.
3. Example.



- i. To create a contingency table between attitude toward health care (dependent or row variable) and gender (independent or column variable) in the present case just enter the numbers.

SDA Tables Program
(Selected Study: 1996 American National Election Study)
Help: [General](#) / [Recoding Variables](#)

REQUIRED Variable names to specify

Row: v960479 **Health insurance**

OPTIONAL Variable names to specify

Column: v960066 **Gender**

Control: _____

Selection Filter(s): _____ *Example: age(18-50) gender(1)*

Weight: v960005a - New Pre-Election Weight **Column percentages**

Percentaging: Column Row Total

Other options

[Statistics](#) [Suppress table](#) [Question Text](#)

[Color coding](#) [Show T-statistic](#)

Figure 3: Cross-tabulation of Attitude on Health Insurance by Gender

- ii. Be sure to check “percentaging.”
- IV. READING A CROSS-CLASSIFICATION TABLE:
- A. The preceding specification produces the table on the next page.
 1. See Figure 4.



		Frequency Distribution		
Cells contain: -Column percent -N of cases		v960066		
		1 Male	2 Female	ROW TOTAL
v960479	1 Government insurance plan	11.8 589	16.6 137	14.4 219
	2 2	9.5 65	11.9 98	10.8 163
	3 3	14.7 101	13.7 113	14.1 214
	4 4	19.1 131	22.5 186	21.0 317
	5 5	18.5 128	14.2 117	16.2 245
	6 6	14.7 102	10.9 90	12.6 191
	7 Private insurance plan	11.7 81	10.2 84	10.9 165
	COL TOTAL	100.0 689	100.0 825	100.0 1,514

11.8 percent of 589 men, 16.6 percent of 825 women favor government health insurance plan.

Figure 4: Attitudes Toward Health Insurance By Gender



- B. Layout
1. The row labels contain the dependent variable number and the value labels.
 - i. Example: the top row labels is "1" for "Government insurance plan."
 - 1) People in this category are (strongly) in favor of public health insurance.
 - ii. Example: the row labeled "4 4" represents those respondents who placed themselves in the fourth category (the one designated by the code "4") on this question.
 2. The column labels have similar meanings and interpretations.
 3. **Marginal totals**
 - i. The "printed" table also contains marginal totals
 - 1) The row marginal totals are the number of people in each category of the dependent variable.
 - a) There are 219 people in the first category of attitude toward government health care.
 - b) Also note that there are 689 men and 825 women.
 4. **Table total**
 - i. This table contains 1,514 "valid" respondents.
 - 1) Note some people have been excluded because they had "missing data" or "invalid" scores on the attitudinal variable.
 - 2) The total number of respondents will vary from table to table because the number of missing data responses varies from question to question.
 5. The inner cells contains both frequencies and **column** percentages.
 - i. We can compare men and women's responses by simply comparing column percentages.
 - 1) Example
 - a) In category 1, "favor a government insurance plan" we find 11.8 percent of men **but** 16.6 percent of women, which amounts to a 4.8 percent difference.
 - 2) In the next row we see that 9.5 of the men and 11.9 of the women placed themselves in the second category or rating of the health care scale.
 - 3) That's not much of a difference.
 - ii. Going down the rows in this fashion we see finally that 11.7 of the men and 10.2 percent of the women put themselves in the "private health insurance category"; again this is not much of a difference.
 6. The frequencies are simply the numbers upon which the percentages are based.
 - i. Example: in the first row, **second** column



$$\left(\frac{137}{825} \right) 100 = 16.6\%$$

- C. Overall interpretation:
1. So, is there a relationship? At first glance it may appear that men and women have roughly the same attitudes.
 - i. True, on the whole women are slightly more apt to choose the government or public end of the scale than private insurance. But the differences are not large.
 2. Still with large contingency tables it's not always easy to make a judgement.
- D. Strategies
1. Data interpretation combines knowledge of simple statistical ideas and lots of substantive judgment. Here are some tools and tips that may assist you.
 2. **Bivariate independent variables:**
 - i. If the independent or column variable has two categories, you can calculate differences of percentages (**not frequencies!**) to determine patterns.
 - ii. Tip: Windows 95 and Windows 98 contain a simple calculator program. Open it to find quickly the percent differences between the two categories.
 - iii. Example:



	Men	Women	Difference
Government insurance	11.8%	16.6%	-4.8
2	9.5	11.9	-2.4
3	14.7	13.7	1.0
4	19.1	22.5	-3.4
5	18.5	14.2	4.3
6	14.7	10.9	3.8
Private insurance	11.7	10.2	1.5
	100% (689)	100% (825)	

- iv. The (percentage) differences are small but note that most of the negative values lie above the scale midpoint, suggesting that women are slightly more inclined to place themselves at the government “side” of the scale whereas men are more likely to locate themselves on the private side.
 - v. The average of the **absolute** value (plus and minus sign ignored) differences is about 3.0. It’s best, however, to look for a pattern as one moves “down” or up a scale, if there is an implied order.
3. **Fourfold table:**
- i. You could make a four-by-four (fourfold) table by “collapsing” the attitude scale into two categories.
 - 1) That is, divide the attitude scale at say roughly the midpoint or category “4.”
 - 2) Example:



	Men	Women
Strong or lean to government health insurance	54.9%	64.7%
Strong or lean to private insurance	45.1	35.3
	100% 689	100% 825

- ii. When examined from this perspective the conclusion that there is a slight relationship between gender and attitudes is confirmed. But it is not strong.
- iii. To recall a previous discussion a “perfect” relationship between X and Y in this type of situation would be

	Men	Women
Strong or lean to government health insurance	0.0%	100.0%
Strong or lean to private insurance	100.0	0.0
	100% 689	100% 825

4. **Color codes**
 - i. The computer program used by SDA has a nice feature: cells are coded according to the expected versus observed proportion or percentages of cases.
 - ii. Although we can't normally see these scheme in action because we usually examine black and white copies of the tables, we could use a shading or grid lines to make the same point.
 - 1) In fact these visual cues are becoming increasingly important in statistical analysis and we may take up the subject latter.
5. Summary statistics.



- i. Most computer programs, including the one SDA uses, calculate summary statistics add additional insights into the relationship between the two variables.
- ii. We'll take up these statistics in the next couple of weeks, but if you want to explore the subject on your own, check the statistics box in the SDA cross-tabulation dialog box.

V. ANOTHER EXAMPLE:

- A. Let's now look at a slightly more complex situation, namely the relationship between attitudes toward defense spending and voting.
 1. In this instance the independent or column variable has several categories while the dependent variable has just three, Clinton, Dole, and Perot.
 2. The variables are:
 - i. Vote for president v961082
 - ii. Attitude toward defense spending (increase or decrease)
 3. Expectation:
 - i. Since Democrats have historically been considered "soft" on defense—that is, generally opposed to large military budgets and willing to cut them or at least maintain them at current levels—and Republicans generally advocated strong national security, a "rational" Democrat would vote for Clinton, while a "rational" Republican would support Dole.
 - ii. What do the data say? Examine the next Figure



Cells contain:
-Column percent
-N of cases

		“Democrat” position				“Republican” position				
		v960463								
		1 Greatly decrease defense spending	2	3	4	5	6	7 Greatly increase defense spending	<i>ROW TOTAL</i>	
v961082	1 Bill Clinton	69.8 24	74.5 73	64.3 105	46.1 140	39.7 85	32.0 33	60.5 22	50.6 481	
	2 Bob Dole	13.1 5	18.2 18	28.4 46	41.6 126	54.4 116	53.5 55	35.0 13	39.8 378	
	3 Ross Perot	10.2 4	4.9 5	5.3 9	10.6 32	5.9 13	11.1 11	4.5 2	7.9 75	
	7 Other	6.9 2	2.4 2	2.0 3	1.6 5	.0 0	3.4 3	.0 0	1.7 17	
	<i>COL TOTAL</i>	<i>100.0</i> 35	<i>100.0</i> 98	<i>100.0</i> 163	<i>100.0</i> 304	<i>100.0</i> 213	<i>100.0</i> 102	<i>100.0</i> 36	<i>100.0</i> 951	

**Includes Perot
and
other**

Figure 5: Rational Voting - Attitudes by Candidate Choice

- iii. We can “ransack” this table like we did the last one. It’s a bit complicated, however, because the dependent variable contains a couple of “extra” categories, categories that may (or may not) hinder our analysis.
 - 1) In particular, not many people respond “other” and the other candidates are not mentioned. So let’s delete that one or treat those responses as missing data.
 - 2) Also, we might want to eliminate Perot votes as well since we are primarily interested in the “two-party” vote and rationality.



- a) This last judgment can be debated. But for now in order to simplify the table let's cut out the Perot and Other categories.:

Cells contain: -Column percent -N of cases		v960463							ROW TOTAL
		1 Greatly decrease defense spending	2	3	4	5	6	7 Greatly increase defense spending	
v961082	1 Bill Clinton	84.2 24	80.4 73	69.3 105	52.6 140	42.2 85	37.4 33	63.4 22	56.0 481
	2 Bob Dole	15.8 5	19.6 18	30.7 46	47.4 126	57.8 116	62.6 55	36.6 13	44.0 378
	COL TOTAL	100.0 29	100.0 91	100.0 151	100.0 267	100.0 200	100.0 88	100.0 35	100.0 859

84.2 percent of the 29 people who want to cut defense spending a lot voted for Clinton. Only 15.8 voted for Dole.

Figure 6: Reduced Cross-tabulation Table

- iv. Except for the last category (i.e., “Greatly increase defense spending”), there is a clear pattern in the table: as support for increased military spending grows—that is, as one moves along the scale from decrease to increase—support for Dole increases.
- v. The relationship is not perfect, but there does seem to be a pattern.
- vi. In fact, if we “lopped off” the last column we could make a strong case that voters are rational in at least this issue area.
- vii. An aside.
 - 1) Actually the situation is more complex. It seems reasonable



to conclude that issue stands are independent and possibly causal variables. So an interpretation based on the following diagram might hold water.

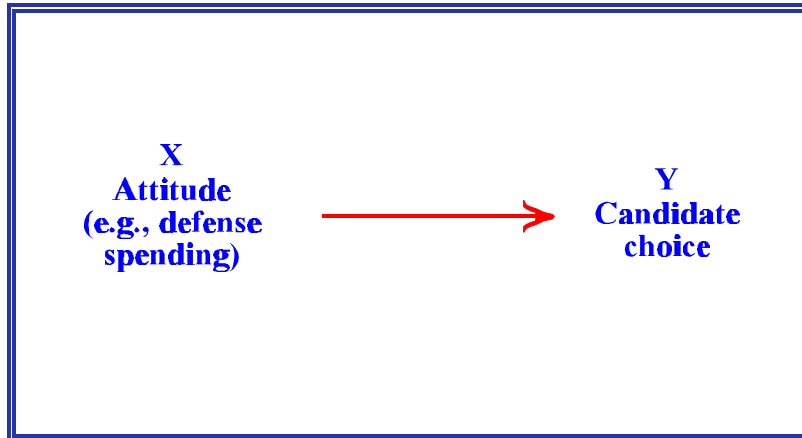


Figure 7: Attitude Causes Behavior

- 2) The arrow is meant to suggest causality.
- 3) But other interpretations are possible as illustrated in the next figure.

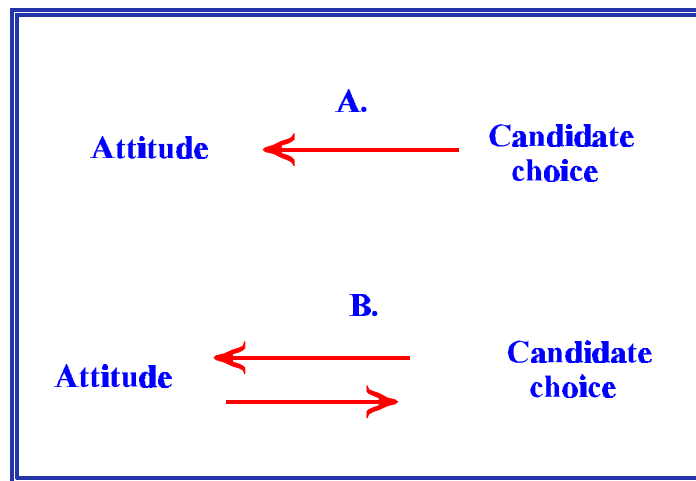


Figure 8: Alternative Interpretations

- viii. Without experimental control and randomization we can't rule out these other possibilities.

VI. SUMMARY:



- A. If time permit we'll look at other examples.

VII. TEST YOURSELF:

- A. Write a one-paragraph interpretation of the following cross-tabulation table.

	Democrats				Republicans		
	Strong	Weak	Leaning	Independent	Lean	Weak	Strong
Clinton	98.4%	89.9%	91.5%	48.9%	25.5%	??%	4.4%
Dole	1.6	??	8.5	51.1	74.5	77.4	95.6
	100% (232)	100% (174)	100% (100)	100% (40)	100% (102)	100% (145)	100% (169)

- 1. You should be able to fill in the blanks (denoted by “??”).
- B. Use the SDA and the 1996 American National Election Study to answer these questions:
 - 1. Is there a racial gap in trust in government (v960566)?
 - 2. Is there a racial gap in attitudes toward abortion (v960566)?
 - i. Does the cynical candidate say one thing about the “right to choose” in a black community and something else in a white constituency?
- C. Note: these are the kinds of questions I am going to ask on assignments and examinations.

VIII. NEXT TIME:

- A. Survey analysis
 - 1. Multi-way tables
 - i. Controlling for other variables.
- B. Reading:
 - 1. Once again see Johnson and Joslyn, *Research Methods*, Chapter 12, pages 327 to 336 describes cross-tabulations.