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

SURVEY ANALYSIS
MULTIDIMENSIONAL TABLES

I.   TODAY'S SESSION:
    A.   Discussion of variables and propositions
    B.   Multidimensional tables
        1.   Statistical approximation to controlled experiments.
        2.   Controlling for "explanatory" and "confounding variables.
        3.   Examples

II.   SUBSTANTIVE PROBLEMS:
    A.   Let's consider two questions.
        1.   What explains participation?
        2.   What role do issues play in candidate selection.

III. PROCEDURE:
    A.   We'll first obtain two-way cross-classifications.
        1.   They show the relationships between Y and X, the dependent and
            independent variables.
    B.   Next we'll introduce a "control" variable.

IV.   RATIONALE:
    A.   "Explication": an attempt to understand or explain a relationship between two
        variables.
        1.   How do we interpret, for instance, a relationship or association between
            the categories of X and Y?
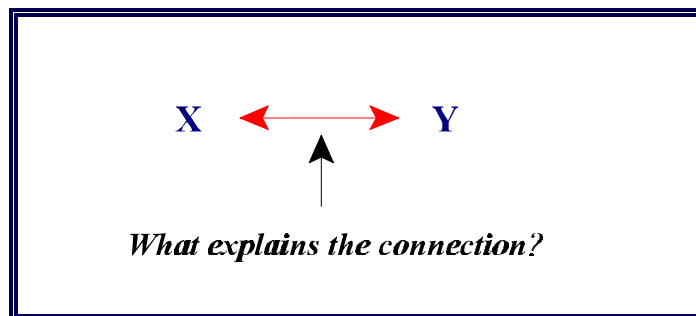
Figure 1: The Problem of Explication

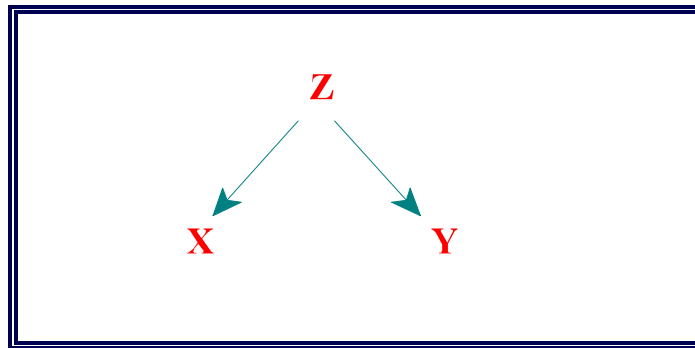B.   We've already discussed one possibility, the spurious relationship.

Figure 2: Spurious Relationship

     1.   The goal of randomization in experimental designs is to check for this possibility.

C.   Another possibility: **specification**

     1.   The relationship between X and Y depends on levels or values of a third variable.
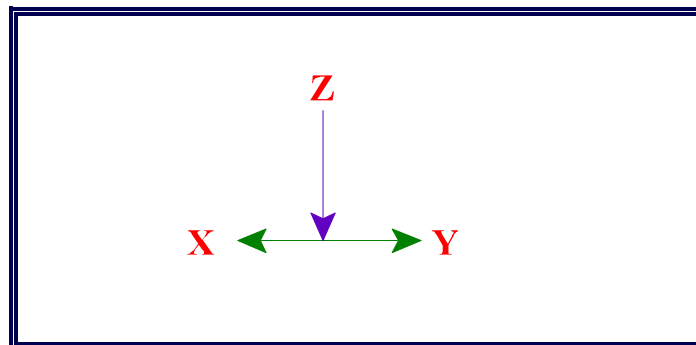

Figure 3: A Contingent Relationship

         i.   The perpendicular arrow is meant to show that the nature–the strength, "direction," and so forth of the X - Y association depends (in part) on the influence of Z, a third variable.

     2.   Verbal example: "there is a connection between issue stands (X) and candidate preference, but the relationship only holds among whites."

         i.   See below.

     3.   It's usually not enough to say that X and Y are related. Instead one must **specify** how they are related. That is, one must indicate what values of X

are associated with what values of Y and under what conditions.
D. To investigate these possibilities or "models" in non-experimental research one has to "introduce" or add one or more control variables to the analysis.
  1. This form of data processing is called **multivariate analysis**.

V. MULTI VARIATE ANALYSIS:
  A. Consider first the connection between attitudes and candidate choice.
    1. Secondary analysis of the 1996 American National Election Study
    2. The dependent variable: who did R vote for in the 1996 presidential election?
    3. The independent variable: R's attitude toward the death penalty.
    4. Some quick theory:
      i. We suspect that those who oppose the death penalty would vote for Clinton, not because **he** opposes capital punishment, but because national Democrats have traditionally been somewhat less supportive of it and have been perceived as being "soft on crime."
      ii. But we also suspect that Clinton enjoys so much support in the black community–this term itself raises lots of issues (is there a "black community?) that non-whites will vote for him no matter what their position on a particular issue happens to be.
      iii. In the language introduced above we might say that **race (Z) specifies the relationship between attitudes (X) and vote (Y)**.
      iv. Or the relationship between attitudes and vote is **contingent** on race.
  B. Method:
    1. We need first the simple cross-tabulation between vote and opinion on capital punishment for all respondents.
      i. Call this the **original** or **total** relationship.
    2. After obtaining it we look at the X versus Y relationship **within** categories of Z, the control variable.
      i. A schematic representation appears in Figure 4 (next page).
      ii. Suppose Z, the control variable, has four categories, labeled 1 through 4.
      iii. Each block represents a cross-tabulation table calculated **only** from the observations that have a particular score on Z.
      iv. We call such an arrangement a "multi-way" or multidimensional table.
      v. In this example suppose that there category 1 of Z contains 40 respondents, category 2 contains 50, category 3 35, and category 4 has 25.
        1) Thus the total N is 40 + 50 + 35 + 25 = 150.
        2) These are **partial** relationships.
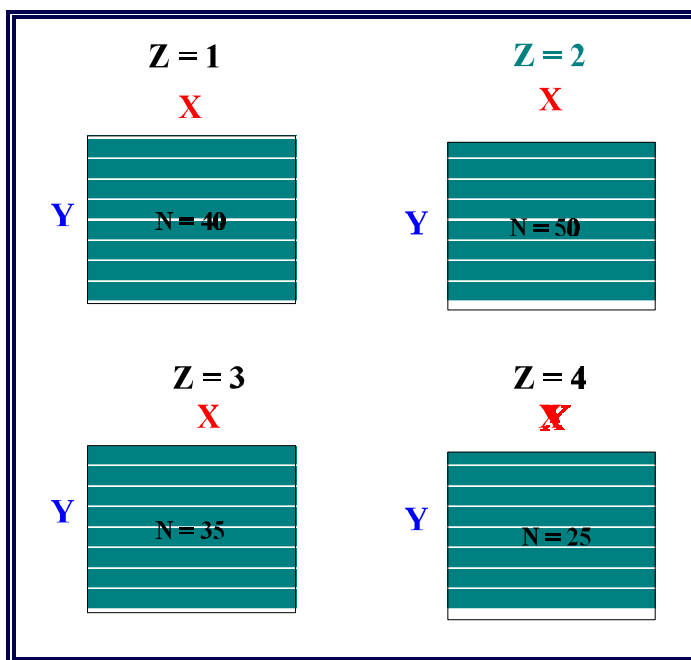
Figure 4: Multidimensional Cross tabulation Table

      vi.   That is, we have (in this case) four "partial" sub-tables. Each one shows the relationship between X and Y within a subset of the data, namely those respondents in the zth category of Z.

VI.   EXAMPLE WITH 1996 ELECTION STUDY:
    A.   We can use the familiar procedures in the SDA system to obtain it.
    B.   First we'll look at the total or original relationship.
        1.   Here's one version of the contingency table produced by the program.
            i.   See Figure 5 on the next page.
            ii.   The data suggest that there is something of a connection between attitudes toward capital punishment and vote for president in 1996. In particular, those who favor the death penalty seem to split their votes between Clinton and Dole–the president received about 3 percent more–with a smattering going to Perot and others.
            iii.   Those opposed to capital punishment, however, voted 3 to 1 for Clinton.
            iv.   We might explore this relationship in more detail by looking at how race affects the situation.

Cells contain: -Column percent -N of cases		v961197		
		1 Favor	5 Oppose	*ROW* *TOTAL*
v961082	1 Bill Clinton	**46.8** 382	**74.7** 155	*52.5* *537*
	2 Bob Dole	**43.1** 352	**19.3** 40	*38.3* *392*
	3 Ross Perot	**8.8** 72	**3.3** 7	*7.7* *79*
	7 Other	**1.2** 10	**2.6** 5	*1.5* *15*
	COL TOTAL	*100.0* *816*	*100.0* *207*	*100.0* *1,023*
	Means	1.68	1.42	1.63
	Std Devs	.87	1.05	.91

Color coding:	<-2.0	<-1.0	<0.0	>0.0	>1.0	>2.0	T
N in each cell:	Smaller than expected			Larger than expected			

Figure 5: Total Relationship Between Vote and Attitude on Capital Punishment

C.   Now we need to add a control variable, race (v960067) to the analysis.
   1.   To do so we simply insert the variable number of the control variable (Z) in the dialog box.
   2.   See Figure 6.

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

REQUIRED Variable names to specify
Row: v961082

Vote (Dependent)

OPTIONAL Variable names to specify
Column: v961197

Opinion (independent)

Control: v960067

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

Weight: v960005a - New Pre-Election Weight ▾

Race (control)

Percentaging: ☑ Column ☐ Row ☐ Total

Other options
☑ Statistics ☐ Suppress table ☐ Question Text
☑ Color coding ☐ Show T-statistic

[Run the Table]   [Clear Fields]

Figure 6: Adding a Control Variable

.

3.   Here we simply "instruct" the program to create a series of cross-tabulations between opinion and vote, one for the respondents in each category of race.

D.   Results:

Statistics for v960067 = 1(White)				
Cells contain:		v961197		
-Column percent		1	5	ROW
-N of cases		Favor	Oppose	TOTAL
	1 Bill Clinton	44.7 / 338	64.5 / 93	47.9 / 431
	2 Bob Dole	45.0 / 340	27.1 / 39	42.1 / 379
v961082	3 Ross Perot	9.1 / 69	4.7 / 7	8.4 / 75
	7 Other	1.2 / 9	3.8 / 5	1.6 / 14
	COL TOTAL	100.0 / 755	100.0 / 145	100.0 / 900
	Means	1.70	1.59	1.68
	Std Devs	.86	1.22	.93

Color coding: <-2.0 | <-1.0 | <0.0 | >0.0 | >1.0 | >2.0 | T
N in each cell: Smaller than expected | Larger than expected

White respondents

Statistics for v960067 = 2(Black)				
Cells contain:		v961197		
-Column percent		1	5	ROW
-N of cases		Favor	Oppose	TOTAL
	1 Bill Clinton	91.0 / 38	100.0 / 57	96.2 / 94
	2 Bob Dole	3.0 / 1	.0 / 0	1.2 / 1
v961082	3 Ross Perot	4.0 / 2	.0 / 0	1.7 / 2
	7 Other	2.0 / 1	.0 / 0	.8 / 1
	COL TOTAL	100.0 / 41	100.0 / 57	100.0 / 98
	Means	1.23	1.00	1.10
	Std Devs	.94	.00	.61

Color coding: <-2.0 | <-1.0 | <0.0 | >0.0 | >1.0 | >2.0 | T
N in each cell: Smaller than expected | Larger than expected
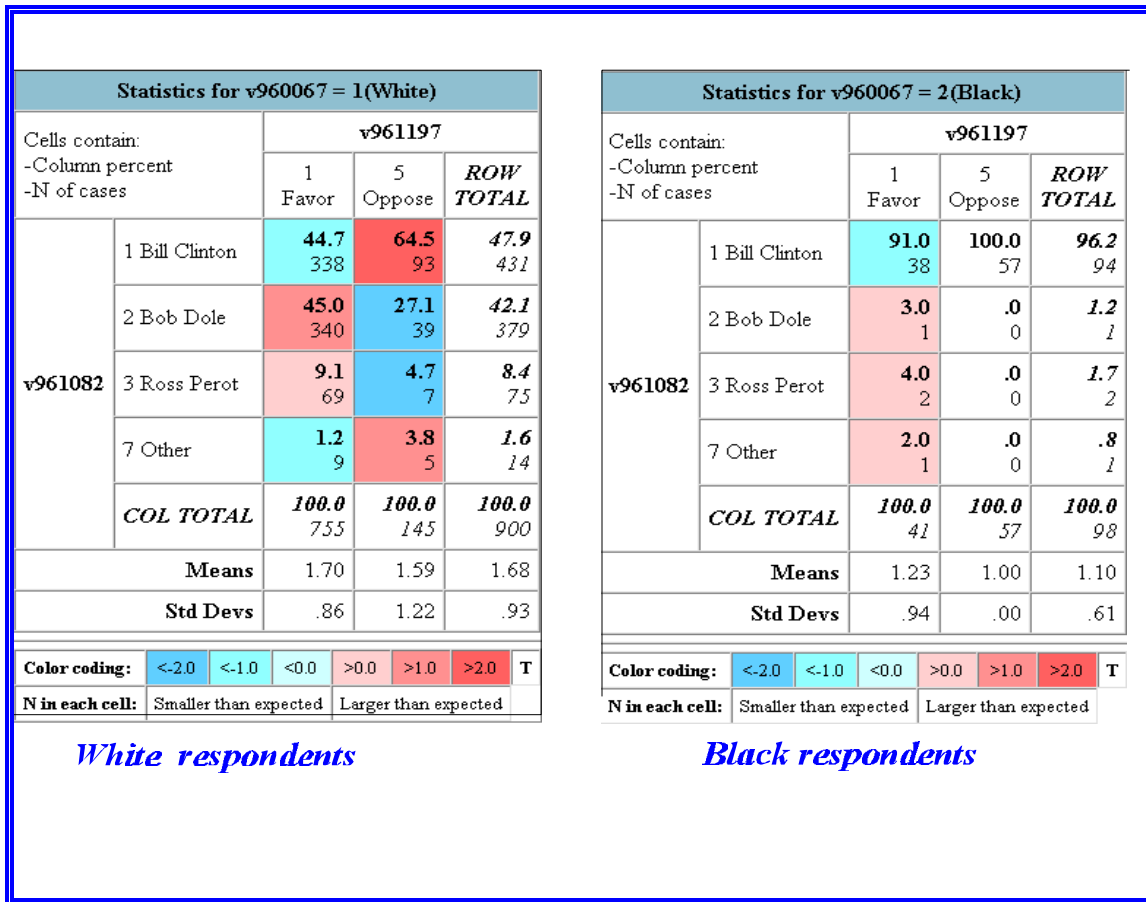
Black respondents

Figure 7: Vote By Opinion by Race

E.   Interpretation

    1.   Look at the first table (on the left). It indicates that about two thirds of the white respondents who oppose capital punishment voted for Clinton, as expected. And, those whites who favor the death penalty split about as indicated before, with a slight plurality (44.7 percent to 45.0 percent favoring the President.

    2.   Now look at the table for black respondents. They are divided: about half favor, half oppose capital punishment. But, almost all, regardless of preference, voted for the President.

    3.   So if there is a relationship between opinion on the death penalty issue and vote, it shows up mainly among white voters.

VII.   ANOTHER EXAMPLE:
   A.   Let's investigate the relationship between turnout (i.e., whether or not a respondent voted, the dependent variable) and interest in the 1996 election campaign (the independent variable) with a control for level of education (Z).
      1.   The variable numbers are
         i.   Voted: v961074
         ii.   Interest v960243
         iii.   Level of education v960610
   B.   We first need the total relationship; that is, the table that shows how turnout and interest are related among all of the (valid) respondents.
      1.   We'll just demonstrate this in class to save space.
   C.   Then we can look at the relationship within educational categories.
      1.   Since the interest variable has 5 categories and the level of education has seven we have a relatively large number of complex tables.
         i.   In particular, since voting has two categories we have seven partial tables (one for each level of education) that are 2 X 5.
         ii.   We need therefore to examine quite a bit of data.
         iii.   In fact there are so many cells to look at making a general interpretation is not easy.

VIII.   RECODING:
   A.   One not entirely satisfactory solution is to reduce the size of the tables and the number of control or partial tables by combing categories.
      1.   That is, instead of using a seven-category educational variable we can combine some categories as follows:
         i.   Categories 1 through 3 = "Low" or "High school or less."
         ii.   Categories 4 through 7 = "High" or "High school or more."
      2.   Similarly we might combine categories of "interest" as follows:
         i.   Categories 1 through 3 = "Interested."
         ii.   Categories 4 through 5 = "Not interested."
   B.   Combining categories like this is called **recoding**.
      1.   When using the SDA system one just enters a variable number and then in parentheses enters the values or codes to be grouped or combined and labels for them.
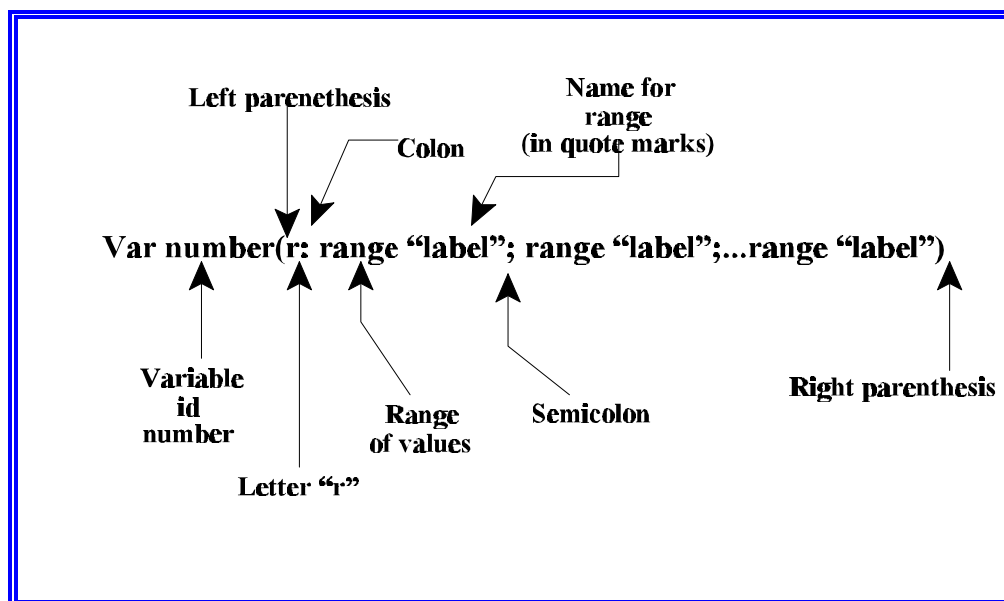   2.   Here's a schematic diagram. Study it carefully because in the SDA system punctuation is critical.

Figure 8: Recode Specification

3. An example: to specify that education is to be recoded in to categories, "low" and "high" we enter in the control variable box:
   i. **v960610(r: 1-3 "Low"; 4-7 "High")**
4. The specification may look complex, but it makes sense.
   i. Type the variable identification number
   ii. In parenthesis type the letter **r** for "recode," the range of values that are to be combined, and then in quote marks a short descriptive label.
5. Another example: to recode interest into "Not interested" and "Interested" categories use:
   i. **v960243(r: 1-3 "Not interested"; 4-5 "Interested")**
   ii. Note the parentheses surrounding the expression, the quotes around the labels; the semicolon separates the recode and label specification.

IX. NEXT TIME:
   A. Additional analysis techniques