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

**MULTIDIMENSIONAL TABLES
(Continued)**

I. TODAY'S SESSION:
    A.   Multidimensional tables
        1.   Mechanics of controlling for "explanatory" and "confounding variables.
        2.   Examples of interpretation
             i.   Recodes
    B.   Important: test yourself

II. RECAP:
    A.   Two problems.
        1.   What explains participation?
        2.   What role do issues play in candidate selection.
    B.   Procedure:
        1.   Obtain two-way cross-classification or **total** relationship
             i.   Interpret this association
        2.   Introduce a "control" variable.
             i.   Divide total pool of respondents into subgroups according to their values on Z, a control variable.
             ii.   This leads to a set of sub-tables.
                  1)   **New notation**
                      a)   Suppose variable Y has I categories and variable X has J categories.
                      b)   If Z has K categories, then there will be K I $\otimes$ J tables where $\otimes$ means "by"
                      c)   Thus "I $\otimes$ J" means "I by J," where as before I is the number of rows and J is the number of columns.
                      d)   Example: Vote (Y) has I = 3 levels or categories, opinion on the death penalty issue has J = 2 categories, and education (Z) has K = 4 categories.
    C.   Purpose:
        1.   Emulate a controlled experiment in which presumably comparison (experimental and control) groups are homogeneous with respect to all background variables or factors.
        2.   "Explication": attempt to understand or explain a relationship between two variables.
             i.   When and under what circumstances is X related to Y.
             ii.   How does the X-Y relationship change with changes in the level of

Z?
3.   Investigation of alternative models.
D.   Method:
1.   Use a statistical program to obtain the original or total relationship.
     i.   Call this the **original** or **total** relationship.
2.   After obtaining it, look at the X versus Y relationship **within** categories of Z, the control variable.
     i.   That is, inspect **partial relationships**.

III.   VOTE BY OPINION BY PARTISANSHIP EXAMPLE:
A.   First look at the total or original relationship.
1.   See the Figure 1.
     i.   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.
     ii.   Those opposed to capital punishment, however, voted 3 to 1 for Clinton.

Cells contain: -Column percent -N of cases		v961197		
		1 Favor	5 Oppose	*ROW* *TOTAL*
	1 Bill Clinton	**46.8** 382	**74.7** 155	*52.5* *537*
	2 Bob Dole	**43.1** 352	**19.3** 40	*38.3* *392*
v961082	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 1: Total Relationship Between Vote and Attitude on Capital Punishment

         iii.   How does race (Z) affect the situation?

B.   Now add the control variable, race (Z) to the analysis.

         1.   To do so we simply insert the variable number of the control variable (Z) in the dialog box.

         2.   Figure 2 shows the program dialog box with entries for the control and independent and dependent variables.

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 2: 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.
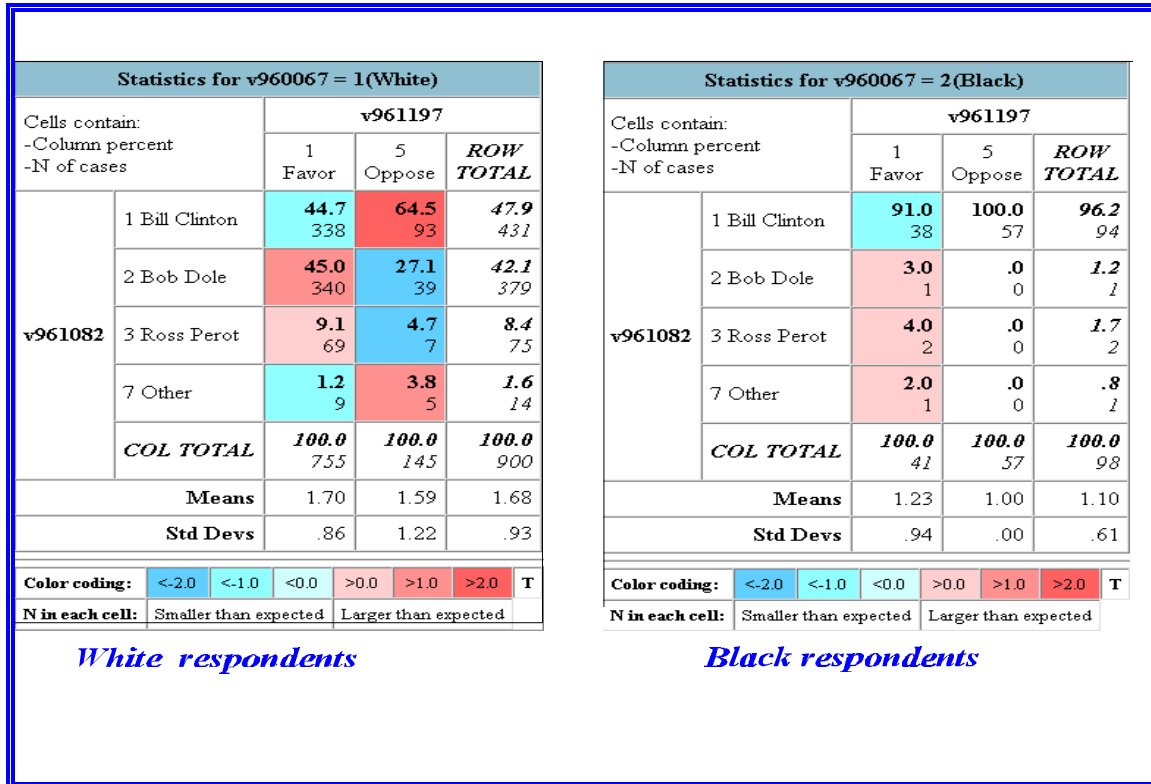
C.   Results:

Figure 3: Vote By Opinion by Race

D.   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.

     i.   This is called a specification: one indicates not simply that X and Y are related but specifies precisely how–under what circumstances and how strongly they are related when a control factor is added to the analysis.

IV.   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.
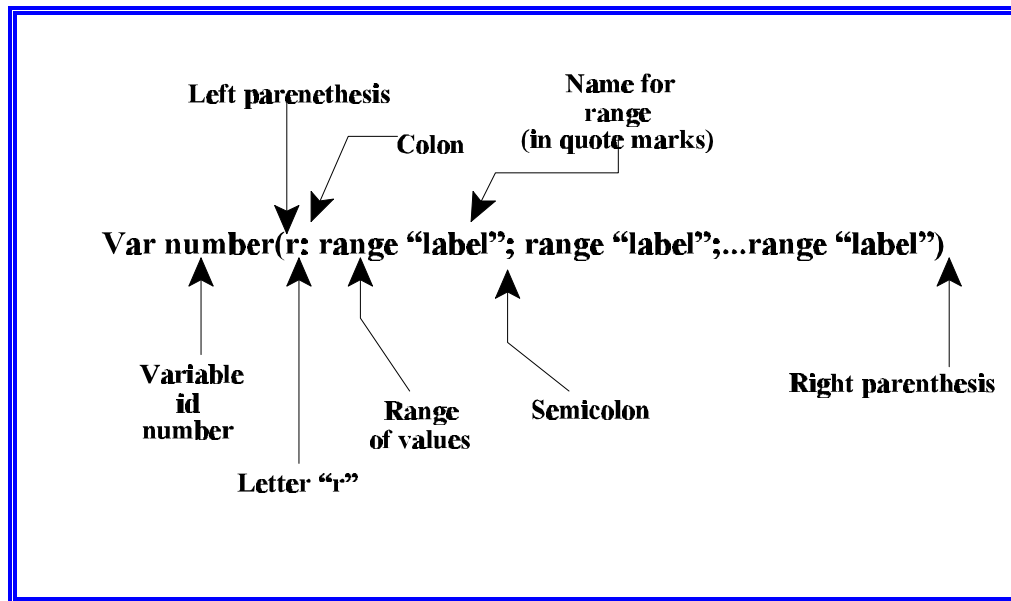

Figure 4: Recode Specification

- 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.

V.   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.   Since the variables interest and education have several categories they perhaps should be recoded.
   1.   That is, since voting has two categories and interest has seven we would have seven partial tables (one for each level of education) that are 2 X 5.
      i.   Keeping with the original numbers of categories would therefore require examining quite a bit of data.
      ii.   In fact there are so many cells to look at making a general interpretation is not easy.
      iii.   The next figure shows how the data can be recoded in the SDA system.

Figure 5: Recode Setup for "Interest" and "Education"

    iv.   Note the parentheses surrounding the expression, the quotes around
           the labels; the semicolon separates the recode and label
           specification.
C.   This specification leads to
      1.   A total relationship table, which incidentally is printed at the bottom of the
            SDA analysis page.
      2.   Two contingency tables, one for Education = "low"(high school and less)
            and one for Education = "high" (more than high school).
      3.   The total relationship appears on the next page.
D.   There appears to be a moderate relationship between the interest and participation:
      there is about a 25 percent difference (81.5% versus 58.8%) in voting.
      1.   Nothing especially exciting here. We would expect this result. What we
            might note, in fact, is that more than half of those in the low interest
            category–the people who express little or no concern with the election on
            1996–still managed to vote. One wonders whether they really did in fact
            vote. Or, perhaps a sense of citizenship compensates for their disinterest.

Cells contain: -Column percent -N of cases		v960243		
		1 interested	2 not interested	*ROW TOTAL*
v961074	1 Yes, voted	**81.5** 772	**55.8** 132	*76.4* *904*
	5 No, didn't vote	**18.5** 175	**44.2** 104	*23.6* *279*
	COL TOTAL	*100.0* *947*	*100.0* *236*	*100.0* *1,183*
	Means	1.74	2.77	1.94
	Std Devs	1.55	1.99	1.70

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 6: Total relationship Between Voting and Interest

2.   Next the **partial relationships**.

Statistics for v960610 = 1(low)

Cells contain: -Column percent -N of cases		v960243		
		1 interested	2 not interested	*ROW TOTAL*
v961074	1 Yes, voted	**72.9** 316	**41.2** 58	*65.1* *374*
	5 No, didn't vote	**27.1** 118	**58.8** 82	*34.9* *200*
	COL TOTAL	*100.0* *434*	*100.0* *140*	*100.0* *574*
	Means	2.09	3.35	2.39
	Std Devs	1.78	1.98	1.91

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			

Statistics for v960610 = 2(high)

Cells contain: -Column percent -N of cases		v960243		
		1 interested	2 not interested	*ROW TOTAL*
v961074	1 Yes, voted	**88.8** 456	**77.2** 74	*87.0* *530*
	5 No, didn't vote	**11.2** 57	**22.8** 22	*13.0* *79*
	COL TOTAL	*100.0* *513*	*100.0* *96*	*100.0* *609*
	Means	1.45	1.91	1.52
	Std Devs	1.26	1.69	1.35

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			

3.   Some remarks:

      i.   Turnout is related to education: people in the high level of education are more likely to have reported voting than those in the lower educational group, no matter what the level of interest.

      ii.   But the difference in participation is greater among the less educated than among the "high" group.

      iii.   Those with low education but categorized "high" on the interest variable "voted 72.9% or 73 percent of the time" whereas only 41.2 percent those in the low interest group said they voted, a 32 percent difference.

      iv.   Among the highly educated, however, the difference among those interested and not interested is 88.8% - 77.2% = 11.6 percent.

      v.   So interest affects participation, especially among those with the least education.

VI.   TEST YOURSELF:

    A.   If you have been following along you should be able to do answer some of these questions or do these assignments.

        1.   Here are two contingency tables showing the relationship between opinions on abortion, gender, and region (i.e., place of residence).

            i.   Suppose the dependent variable is attitudes toward abortion as measured by answers to this question: "Are you for or against legalized abortion?"

                 1)   What are the independent and control variables?

	North				South	
	Male	Female			Male	Female
For	55%	85%	For		35%	40%
Against	45	15	Against		65	60
Total	100% (120)	100% (150)	Total		100% (75)	100% (50)

            ii.   Could you if your life depended on it, construct the total table given this information?

            iii.   In words, what is the relationship between gender and opinions on legalized abortion among southerners? Among northerners?

            iv.   By the way, is the question wording in this example very good?

        2.   Last time we talked about "teasing out" the meaning of verbal statements. I claimed that standardized tests for professional schools are full of these

kinds of problems. Here is an example form the "Registration and Information Book" of the Law School Admission Test. I may be crazy, but this question and all the others like it have the same structure as the questions I asked about the Democratic and Republican platforms.

i.   "Some legislators refuse to commit public funds for new scientific research if they cannot be assured that the research will contribute to the public welfare. Such a position ignores the lessons of experience. Many important contributions to the public welfare that resulted from scientific research were never predicted as potential outcomes of that research. Suppose that a scientist in the early twentieth century had applied for public funds to study molds: who would have predicted that such research would lead to the discovery of antibiotics–one of the greatest contributions ever made to the public welfare?

   1)   "Which one of the following most accurately expresses the main point of the argument?

   a)   The committal of public funds for new scientific research will ensure that the public welfare will be enhanced.

   b)   If it were possible to predict the general outcome of a new scientific research effort, then legislators would not refuse to commit public funds for that effort.

   c)   Scientific discoveries that have contributed to the public welfare would have occurred sooner if public funds had been committed to the research that generated those discoveries.

   d)   In order to ensure that scientific research is directed toward contributing to the public welfare, legislators must commit public funds to new scientific research.

   e)   Lack of guarantees that new scientific research will contribute to the public welfare is not sufficient reason for legislators to refuse to commit public funds to new scientific research.

3.   Using the SDA archive's 1996 National Election Study data, construct the total relationship between respondents' assessments of Clinton's moral character (v960423) or his honesty (v960429) and gender and then see if this relationship (if any) changes when party identification (v960420) is introduced as a control.

VII.   NEXT TIME:
    A.   More examples
    B.   How to present tables in written documents
    C.   Measurement
    D.   Reading:
        1.   As noted before, Johnson and Joslyn, *Research Methods* pages 327 to 336 discuss simple 2 by 2 tables.
        2.   Read pages 378 to 386 for multivariate crosstabulation.