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:
      2. Cross-tabulations
         i. We need information of the following sort.

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.

   ![Cross-tabulation Dialog Box](image)

   **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 paste 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.

![Figure 3: Cross-tabulation of Attitude on Health Insurance by Gender](image)

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

<table>
<thead>
<tr>
<th>Cells contain:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Column percent</td>
</tr>
<tr>
<td>- N of cases</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1 Male</th>
<th>2 Female</th>
<th>ROW TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Government insurance plan</strong></td>
<td>11.8</td>
<td>16.6</td>
<td>14.4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>137</td>
<td>219</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>9.5</td>
<td>11.9</td>
<td>10.8</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>98</td>
<td>163</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>14.7</td>
<td>13.7</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>101</td>
<td>113</td>
<td>214</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>19.1</td>
<td>22.5</td>
<td>21.0</td>
</tr>
<tr>
<td></td>
<td>131</td>
<td>186</td>
<td>317</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>18.5</td>
<td>14.2</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>128</td>
<td>117</td>
<td>245</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>14.7</td>
<td>10.9</td>
<td>12.6</td>
</tr>
<tr>
<td></td>
<td>102</td>
<td>90</td>
<td>192</td>
</tr>
<tr>
<td><strong>7 Private insurance plan</strong></td>
<td>11.7</td>
<td>10.2</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>81</td>
<td>84</td>
<td>165</td>
</tr>
<tr>
<td><strong>COL TOTAL</strong></td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td>689</td>
<td>825</td>
<td>1,514</td>
</tr>
</tbody>
</table>

Figure 4: Attitudes Toward Health Insurance By Gender

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

\[
\left( \frac{137}{825} \right) \times 100 = 16.6\%
\]
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:
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

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong or lean to government health insurance</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Strong or lean to private insurance</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>689</td>
<td>825</td>
</tr>
</tbody>
</table>

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
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.
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.

<table>
<thead>
<tr>
<th></th>
<th>Democrats</th>
<th></th>
<th>Republicans</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong</td>
<td>Weak</td>
<td>Leaning</td>
<td>Independent</td>
<td>Lean</td>
<td>Weak</td>
</tr>
<tr>
<td>Clinton</td>
<td>98.4%</td>
<td>89.9%</td>
<td>91.5%</td>
<td>48.9%</td>
<td>25.5%</td>
<td>??%</td>
</tr>
<tr>
<td>Dole</td>
<td>1.6</td>
<td>??</td>
<td>8.5</td>
<td>51.1</td>
<td>74.5</td>
<td>77.4</td>
</tr>
<tr>
<td></td>
<td>100% (232)</td>
<td>100% (174)</td>
<td>100% (100)</td>
<td>100% (40)</td>
<td>100% (102)</td>
<td>100% (145)</td>
</tr>
</tbody>
</table>

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.