DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Research Methods Posc 302

RESEARCH DESIGN

I. TODAY'S SESSION:
- A. Three types of empirical research
  - 1. Sample surveys

II. SUBSTANTIVE EXAMPLES:
- A. Questions in turnout:
  - 1. Comparison of partisans and non-partisans
  - 2. Turnout of ethnic groups with similar social and economic backgrounds
  - 3. Trends in turnout.
- B. We can used secondary analysis of sample surveys to illustrate research designs appropriate for investigating these questions.

III. SAMPLE SURVEY:
- A. "Direct" solicitation of information from (sample of) subjects.
- B. Elements:
  - 1. Questionnaire (i.e., measurement and operational definitions)
  - 2. Population definition
  - 3. Sample design
  - 4. Data management
  - 5. Analysis

IV. MEASUREMENT:
- A. Turnout:
  - 1. The **Dependent variable**
    - i. Variable because it varies among individuals
    - ii. Dependent because it depends, we suppose, on other factors such as age, education, and ethnicity.
  - 2. Measure:
    - i. Reported self- voting in last election or in past elections.
    - ii. A **binary** or **dichotomous** variable:
      - 1) It has just two values or classes: voted and did not vote.
    - iii. A problem: respondents over report participation.
      - 1) Surveys report turnout rates that are 20 to 30 percent too high.
- B. Partisanship
  - 1. An **independent variable**.

i. "Independent" because we do not here try to "explain" it variation.
ii. A variable because some people are very partisan, some moderately so, and some not at all.

2. Measure: self-reported degree of identification with one of the major political parties.
   i. **Ordinal** variable:
      1) Values of the variable, responses classes, have an implicit order or magnitude.
         a) Many researchers assume an underlying numerical or quantitative scale that is not observed.
      2) Scale: strong to none (non-partisan)
      3) Note: although a numerical measurement may be implied, it is not observed and we can't say that a strong partisan is twice as committed to a party as a plain partisan.

3. Ethnicity
   i. An independent variable because it does not depend on other factors in our study.
   ii. Measure: respondent self placement or identification.
      1) **Categorical** or **nominal** variable
      2) Values of the variables are simply labels or names or categories.
         a) There is no implicit numerical order among them.
         b) High and low values are meaningless.
         c) We can't make numerical comparisons.

4. Trend
   i. An **numerical** independent variable
      1) Numerical because we will simply use a time counter.
   ii. Time is indicated by **year of survey**
      1) The difference between 1996 and 1998 is two years as is the difference between 1966 and 1968.
      2) We can speak meaningfully of the growth in turnout in a 10 year period.

V. POPULATION AND SAMPLE:
   A. Population; well defined collection of objects (units of analysis)
   B. Sample: subset of objects selected in some fashion.
      1. Simple random sample: members of sample have an equal probability of being included in the sample.
      2. Non-simple sample: some members have different or unknown chance of being included.
   C. We want to be able to generalize about the adult (voting age) population of the United States.

D. We'll discuss sampling procedures in more detail later, so for now we'll assume that the respondents to the survey constitute a random sample of non-institutionalized American citizens 18 years of age or older who reside in the continental United States.

VI. DATA MANAGEMENT:

A. Questionnaires are "coded" and information is stored in a medium suitable for electronic (computer) data processing.

B. Note: in survey research variables are often questions and the values of the variables are the possible responses.

C. **Codebook**:

1. List questions and other information obtained during the course of the interview.
   i. Question are usually reported verbatim along with instructions to interviewer.
   ii. Example:

Generally speaking, do you usually think of yourself as a Republican, a Democrat, an Independent, or what? (IF REPUBLICAN OR DEMOCRAT) Would you call yourself a strong (REP/DEM) or a not very strong (REP/DEM)? (IF INDEPENDENT, OTHER [1966-1992: OR NO PREFERENCE]:) Do you think of yourself as closer to the Republican or Democratic party?

Box 1: Sample Codebook Entry

2. List responses to questions.
3. States numerical or non-numerical codes assigned to each response.
4. Often gives raw frequencies or the number of respondents in each class or value of the variable.
5. Data management information
   i. List variable names and **locations** on the storage medium.
   ii. Data processing information.
6. See the appendix to these notes for a complete example

D. A data file must be "processed" and analyzed before its contents can be interpreted.

1. Look at the codebook entry for partisanship at the end of the notes.
   i. The question pertains to party identification with strong Democrats separated from strong Republicans
   ii. But we want just "strong" partisans, so before analyzing and interpreting data we will want to recode this variable as follows:

1) Strong Republicans and strong Democrats are classified strong partisan coded 1, for example.
2) Moderate Democrats and Republicans are coded just "partisan" and scored, say, 2.
3) Independents who "lean" to one of the parties are "weak" partisans and coded 3.
4) Finally, those who are independent are coded 0 for "non-partisan."
5) The new variable is thus:

```
      LEVEL OF PARTISANSHIP
1. Strong partisans
2. Partisans
3. Weak partisans
4. Non-partisans (independents)
```

Box 2: Constructed Variable

6) We'll illustrate the procedure many times in the coming weeks.

VII. ANALYSIS:

A. Techniques, frequency distribution, cross-tabulation table, line plot, statistical test of significance:

1. Simple frequency distributions
   i. How many people gave each response or fell into or were assigned to each category or value.
2. Cross-tabulation:
   i. A joint frequency distribution.
   ii. How many respondents in a particular category (value) of a variable are in the categories of another variable.
3. Example next page.

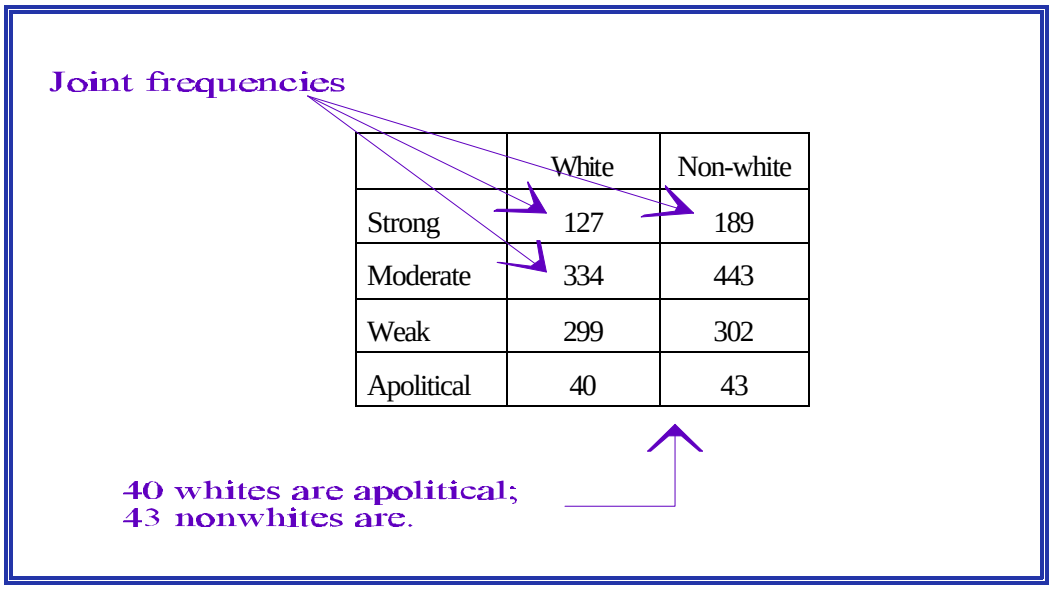


Figure 1: Cross-tabulation

This hypothetical example demonstrates one variable ("ethnicity") cross classified or cross tabulated with another, "degree of partisanship."

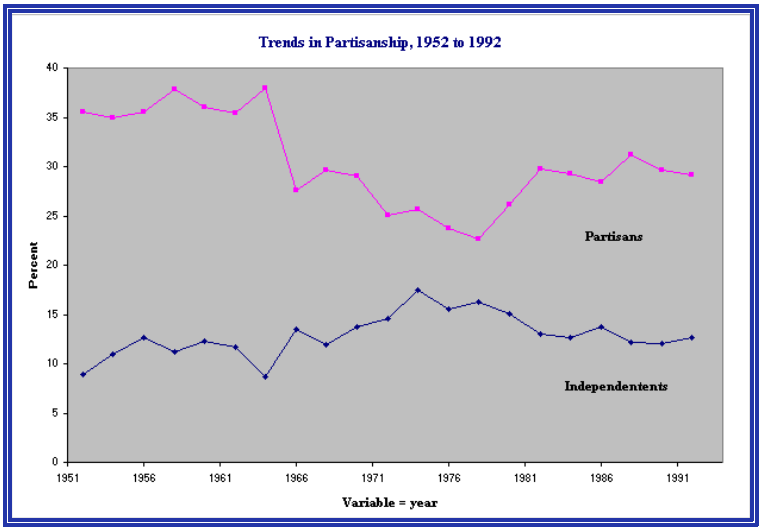
i. But itself the tables provides only a limited amount of information, but it can be easily supplemented with additional material.

4. A cross-classification is an elementary way to build a model that shows how one variable or factor is related to another.

B. Plots

1. When displaying and interpreting "time-series" data–information collected on units of analysis at different time periods, we'll use time as one variable.

i. Here's an example.



2. We'll discuss these plots later.

C. Statistical inference:

1. Since we just have a sample from a population, we need to make guesses or inferences about the population.
   i. Our sample may show a relationship between partisanship.
   ii. But is this observed observation the result of sampling error or does it point to the existence of a "real" relationship.
2. Again, more later.

VIII. ANATOMY OF A TWO-WAY CROSS-CLASSIFICATION TABLE:

A. General features

1. A (two-way) cross-classification table, also called a cross-tabulation or "cross tabs" for short or contingency table, tabulates the number of occurrences of combinations of categories of two variables.
   i. The term "two way" arises because two variables are cross-tabulated.
   ii. In a three-way table three variables are cross-classified.

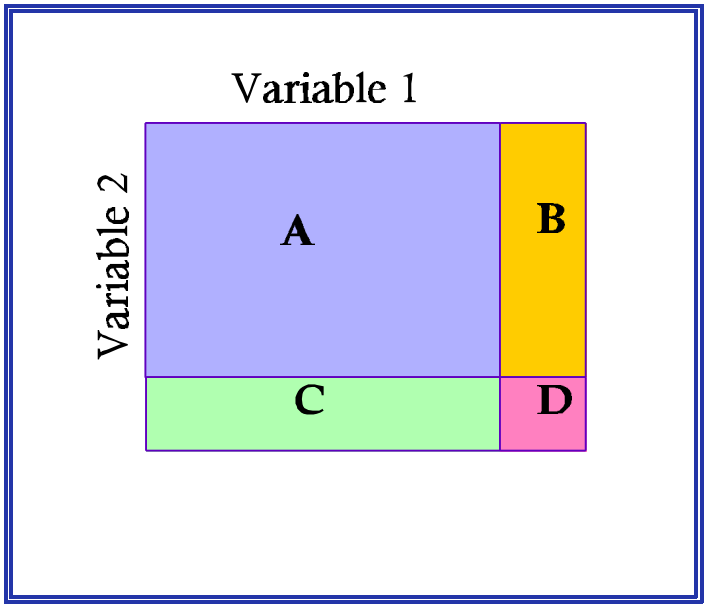


Figure 3: Parts of a Cross tabulation Table

2. **A** body of the table contains **cell** frequencies: the number (and percent) of cases or observations in the combined values of the variables.
3. **B** Row marginal totals: the number of cases in each row of the table, which is the number of cases in each category of variable 2.

4. **C** Column marginal totals: the number of cases in each column of the table, which is the number of cases in each category of variable 1.
5. **D** The table total or total number of cases or observations in the table.

IX. DETAILED EXAMPLE:

A. Example of cross classification between partisanship, the independent or column or horizontal variable, and turnout, the dependent or row or vertical variable.

B. Aspects

1. Layout and titles
2. Variable names
3. Marginal frequencies
   i. The number of cases in each category of the variables.
4. Table total **N**, the number of "cases" in the table.
5. Cell frequencies: the number of cases having value i on the dependent variable ***and*** value j on the independent variable.
   i. Example: 416 are code 1 on partisanship ***and*** 1 on voted.
   ii. Similarly, 366 are coded 2 on partisanship ***and*** 1 on voted.
   iii. The entries are just the number of occurrences of each combination of variable values.
6. Percentages: in this table cell entries are percentages, which the number in a category or combination of categories divided by a total and multiplied by 100.

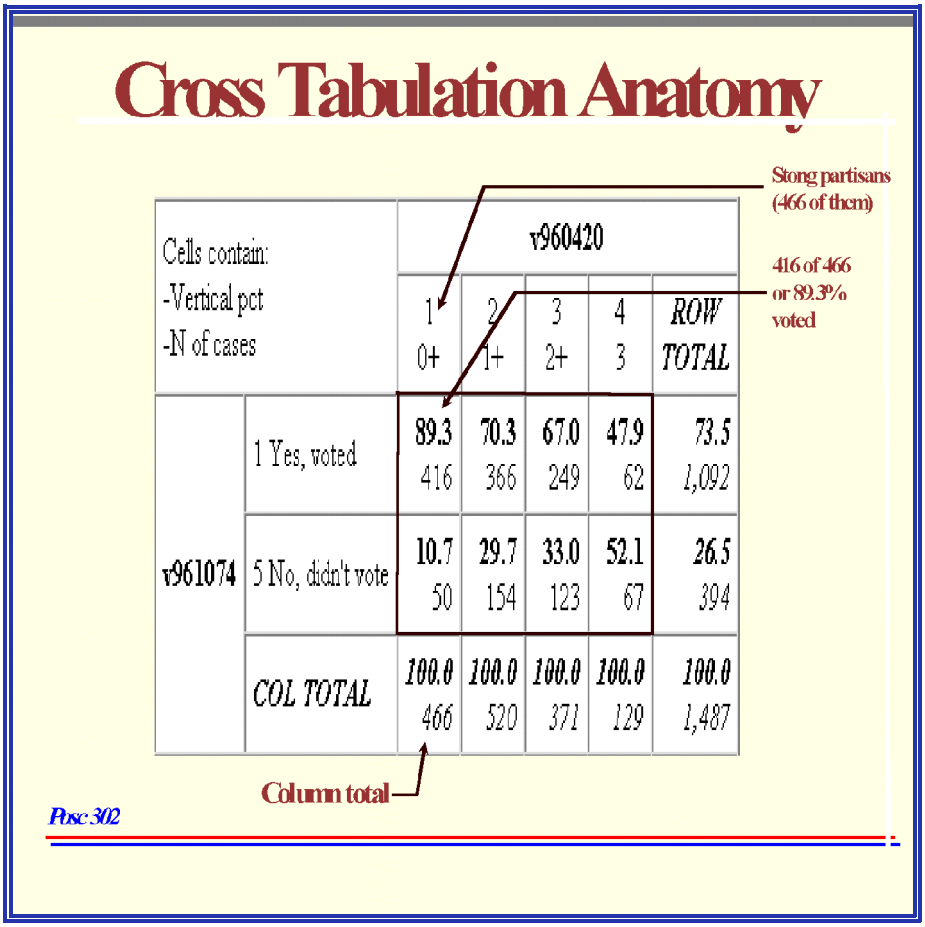


Figure 4: Column Percentages

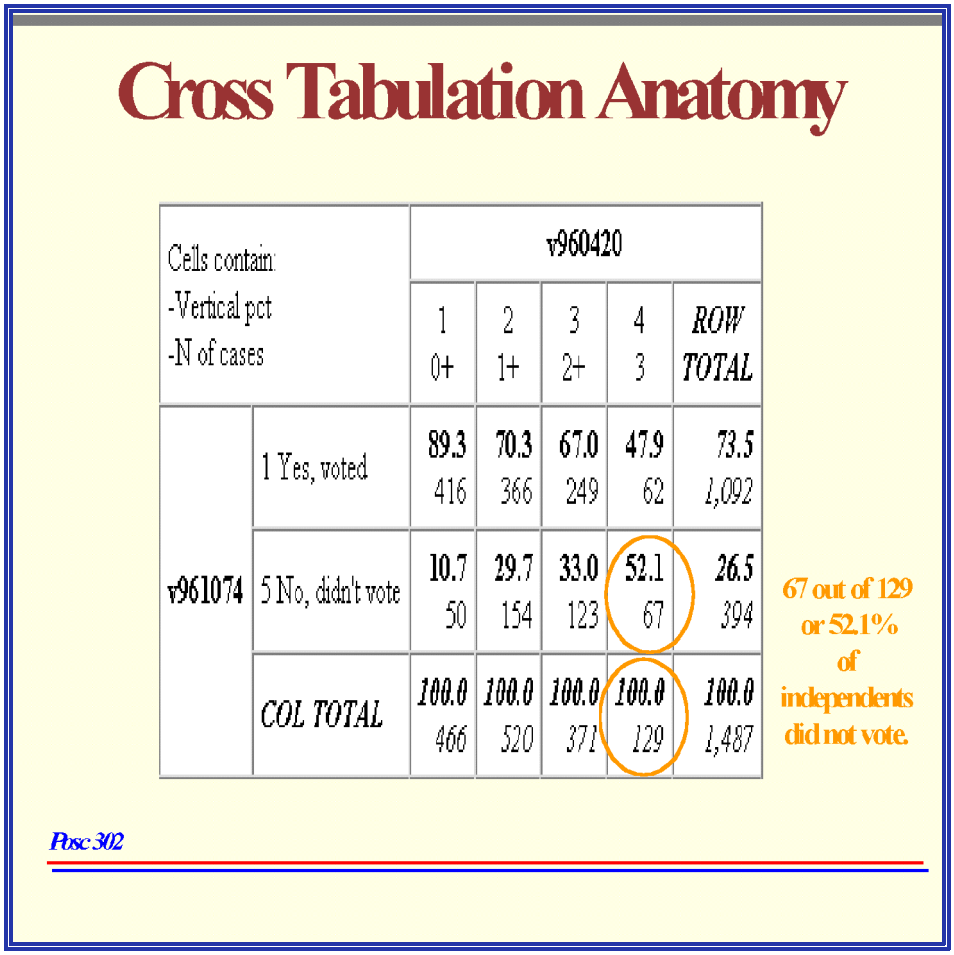
i. Percentage:

$$Percent = \left(\frac{Cell\ frequency}{category\ total}\right)100 = \left(\frac{N_{ij}}{N_{+j}}\right)100$$

Percentage calculation

ii. For now forget about the notation.

7. In this example we see that about 89.3 percent of the strong partisan respondents report having voted.
8. Example:



i. Here we see that 67 out of 129 independents said they didn't vote. That's 52.1 percent.

C. KEEP THE INTERPRETATION OF PERCENTAGES STRAIGHT!

X. INTERPRETATION:

A. Using percentages look for patterns.

1. In the previous case we can make the generalization:
   i. In this sample as partisanship declines, turnout goes down.
2. A restatement:
   i. Partisans are more likely to vote than are independents or weak partisans.

XI. NEXT TIME:

A. Experiments and quasi-experiments

B. Reading:

1. Johnson and Joslyn, ***Research Methods***, Chapter 5.

```
Partisanship 1996
SUMMARY - R PARTY ID
-----------------------------------------------------------------------

See K1, K1a/b, K1c for full question texts.
Built from K1, K1a/b, K1c

Code 8 (apolitical) was used only if the respondent had a code of 5 (no
preference) in K1 and a code of 3,8 or 9 in K1c and also showed little or
no interest in politics in response to the following questions: A1 (pre),
B1 (pre), N2 (pre), F1 (post).

    PCT     PCT      N  VALUE  LABEL
  VALID     ALL
   19.4    19.2    329      0  Strong Democrat (1,1,0 in K1, K1a/b, K1c)
   19.7    19.5    334      1  Weak Democrat (1,5/8/9,0 in K1, K1a/b, K1c)
   13.7    13.6    233      2  Independent-Democrat (3/4/5,0,5 in K1,  K1a/b, K1c
    8.5     8.5    145      3  Independent-Independent (3,0,3/8/9 in K1, K1a/b, K1c;
                              5,0,3/8/9 if not apolitical)
   10.8    10.7    183      4  Independent-Republican (3/4/5,0,1 in K1, K1a/b, K1c)
   15.2    15.0    257      5  Weak Republican (2,5/8/9,0 in K1, K1a/b, K1c)
   12.6    12.5    214      6  Strong Republican (2,1,0 in K1, K1a/b, K1c)
    0.1     0.1      1      7  Other; minor party; refuses to say (4,0,3/8/9 in K1, K1a/b,
                              K1c)
            0.8     14      8  Apolitical (5,0,3/8/9 in K1, K1a/b, K1c and no interest in
                              politics)
            0.2      4      9  NA (8/9,0,0 in K1, K1a/b, K1c)
                ----
                1714 cases

    Type: numeric     Min: 0     MD Codes: 8,9-*
    Decimals:   0     Max: 7
```

Box 3 Codebook Entry