DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Research Methods Posc 302

SECONDARY ANALYSIS OF DATA
Final Projects

I. TODAY'S SESSION:
- A. Secondary data analysis
- B. Research designs for final projects
- C. Importance and statistical significance.
  1. See the attached News Journal article and the remarks in the notes.
- D. Writing tips:
  1. Be precise and crisp.
     - i. Example:
       1) "**...95 percent voted for Dole."**
       2) Not, "...95 percent voted in favor of Dole."
     - ii. Example:
       1) "...researchers asked **the** subjects several questions **about** their beliefs."
       2) Not "...researchers asked **their** subjects several questions **relations to** their beliefs."
          a) This last error (and it's an error) could have been caught if the person had proofread out loud.
  2. The next time I see "**e**ffect" for "**a**ffect" or vice versa I'm going to limit the grade to C- at a maximum no matter how wonderful the work is.
  3. Manually proofread. A spell checker will not pick up "then" that is used for "than."
     - i. More than one or two blatant typos of this sort will put an A out of your reach.

II. SECONDARY ANALYSIS:
- A. Substantive theory or propositions or ideas or claims.
  1. The final project asks you to think about American public opinion. Subject to the limitations stated in the assignment you can investigate various questions about the public's political beliefs, attitudes, and behavior.
     - i. Your grade rests partly on how skillfully and imaginatively you frame the topic.
- B. As you do so you will of course establish some abstract concepts such as generation gap, political culture, ideology, and the like.
  1. The problem is that you can't then design your own survey or poll to "measure" the abstractions but must instead rely on what has already been included on publically available surveys.

2. This reliance on "secondary data" leads to the slippage or gap between a concept and its empirical indicators that we discussed in a previous class.
3. In fact, your definition of say political ideology may have nothing to do with the sort of questions the surveys include.

C. So, what is to be done?
1. First think about the concepts.
2. Look for items that might be reasonable measures of them.
3. If none exist, reformulate or rethink the definition of the concept(s)
4. Look again for empirical indicators.
   i. You should have roughly 6 to 10 questions in mind.

D. When choosing these variables think a little about how much data you can analyze and interpret.
1. Suppose you have one independent variable (e.g., age groups), six dependent variables, and two control variables.
2. You would then have six two-way tables that describe the original relationships and two sets of control tables.
   i. Consider a more specific example: X has 2 categories and each Y has 5.
   ii. Then you have 6 5 by 2 tables.
   iii. Now suppose a control variable has 3 categories. It will thus generate six 5 by 2 by 3 tables.
3. This may or may not be too much to report. We'll discuss ways to analyze and present the data efficiently.

E. Moreover, don't just pick them at random. Think about why you want to include each variable.

III. PARTICULAR PROJECTS:

A. My project pertains to social class differences in political attitudes and behavior.

B. As indicated in the last set of notes, I could outline the research design as follows.
1. Social Class Polarization In the United States
   i. I wonder if the lower class differs significantly (in a theoretical sense) from the upper class in terms of economic outlook.
      1) Although the economy continues to boom, not everyone shares in the growth equally. In fact, lots of Americans may be "falling" behind.
      2) Some economists argue that the middle class is shrinking as both ends of the income/wealth distribution grow larger.
      3) A simplistic statement is that the United States may be moving in the direction of a nation of haves and have nots.
      4) There is also evidence that the labor market has been fragmented into "primary" and "secondary" sectors.
         a) The primary sector consists of jobs in industries that

are growing. These jobs are characterized by relatively high wages, security, benefits, chances for promotion, and prestige.

b) The secondary market consists of marginal occupations requiring low skills and offering lower wages and benefits.

c) One wonders if these changes affect political attitudes.

5) A related concern is false consciousness: does the lower class realize that its economic well being has in fact not improved much in the last 20 years and may have even slipped?

ii. To answer some of these questions I want to compare members of different classes along a number of attitudinal dimensions.

2. Data: 1996 American National Election Study
3. Independent variable: I am going to use occupation (re-coded lightly) as the indicator of social class.
   i. The particular item is: v960676 "Pre. Stacked -- R collapsed occupation code"
4. Dependent variables: I am going to investigate class differences on respondents' perceptions of the economy. The main variables are:
   i. v960338 "Pre. Is R much better or worse off financially than a year ago"
   ii. v960340 "Pre. Does R think R will be much better/worse off financially next year."
   iii. v960386 "Pre. R think economy has gotten much better/worse over past year"
   iv. v960388 "Pre. R expects economy to get much better or worse over the next year"
   v. v960389 "Pre. R think the standard of living will be better or worse in 20 years"
   vi. v960391 "Pre. Have the federal government policies made nation's economy much better or worse"
5. Control variables: to make sure that any differences between classes are not due just to education or family income I will control for these two variables.
   i. v960610 "Pre. Summary of R's education"
   ii. v960701 "Pre. R's family income in 1995"
   iii. I realize that I will probably have to combine some categories in these control variables to reduce the number of tables.
6. If I don't find a relationship between class (as measured by occupation) and some of the dependent variables, I will discard those items. But if any of

the dependent variables are related to class, I'll then further investigate by controlling for family income and education. I am especially interested in knowing if the lower class is more or less or equally optimistic about its and the nation's economic future.

C. Is there a racial gap in the United States?
   1. The idea behind this question should be fairly obvious.
      i. Because it is a straight forward I want an honest effort to locate some background "literature" that discusses the question.
   2. Moreover the independent variable, race or ethnicity, is pretty clearly described by questions available in the National Election Study or the General Social Survey.
   3. What is necessary is to
      i. **Define** gap. What is a substantively significant gap?
      ii. What items (attitudes, beliefs, etc.) will best illustrate the presence or absence of a gap?
         1) Clearly questions about affirmative action and race relations come to mind.
         2) But these have been studied to death.
         3) Why not use variables that pertain to politics in general. You could, for example, use items that refer to social issues. Or to ones that deal with non-race related public policy matters such as spending on defense and foreign aid.
            a) One of my colleagues suggested to me that African Americans are more conservative (or just as conservative) on social issues such as legalization of drugs as are whites. I never got around to investigating the matter. I would find a paper on this subject immensely informative.
   4. After establishing that there are gaps on certain issues you should ask if these relationships will disappear when social and economic status factors are controlled.
      i. Consider this argument: sure there's differences between whites and blacks on issue X. But that's only because positions on that issue are determined by income. People in the lower classes tend to be for X while those in the upper class oppose it. And since income also tends to be related to race, the relationship will be spurious.
      ii. You can address this claim by explicitly controlling for income to see if the original relationship vanishes or is at least weakened, as was demonstrated in a different context in Assignment 4.

D. Is there a generation gap?
   1. The key variable is generation.
      i. Note: the variable is **generation**, not generation gap. (Make this

mistake at your peril.)

ii. How can it be operationalized. Generations are defined as people who are born or come of age in particular periods.
   1) The baby boom generation members, for instance, were born between 1949 and (say) 1960. Post-Cold War generation, you, was born starting about twenty years ago.

iii. So the first step in this project is to define generation.
   1) A search of the Internet with the tools and methods discussed earlier in the semester should lead you to some definitions.

iv. Essentially, figure out when people at the upper and lower age limits of the generations were born and then figure out how old they would be in 1996 when the 1996 study was conducted.

v. These age limits will then indicate how the variable age should be recoded.
   1) We may see some examples in class.

2. As with the race gap, the next step is to ask on what issues would the generations be likely to differ. I would imagine social issue such as abortion and other life style questions would be a possibility. But a common argument is that the "younger generation" (whatever that means) is more "fiscally" conservative and more "socially" liberal than older people. This too could be the basis of a great paper.
   i. A's (of course) go to those who besides showing research skills display some imagination and thought in refining this topic.

3. And as with some of the other projects it might be appropriate to control for social and economic factors.
   i. Young individuals presumably earn less than middle aged people. And perhaps money makes people conservative. So, if one looked at a group of citizens all having roughly the same income, the age factor might disappear. I don't know. That's why I am very curious to see.

IV. SUBSTANTIVE VERSUS STATISTICAL SIGNIFICANCE:

A. For your information.

B. When a someone rejects a null hypothesis in favor of an alternative the result is said to be **statistically significant**.

1. Example: if we reject the hypothesis of (statistical) independence between X and Y because the observed chi square occurs with probability of .05 or less, we say that the variables are significantly related.
   i. Frequently, the phrase "at the .05 level" is added so the sentence becomes "There is a significant relationship between X and Y at the .05 level."

C. But the term or phrase "significantly related" means only that we have rejected a statistical null hypothesis, namely that X and Y are independent.

D. The relationship may be **substantively** or **theoretically** unimportant or perhaps weak.

1. In other words, rejecting a statistical hypothesis does not mean we have found something of great interest or importance. We may have, but may be not.

E. Here's an example:

1. Suppose we hypothesize that X and Y are statistically independent and collect data to test this proposition.
2. We agree to use the .05 level. That is, we'll calculate a chi square for the observed table based on N observations, and if the probability of getting this result is .05 or less, we'll reject the null hypothesis of independence.
3. Suppose the observed table is the one shown on the next page.
4. Most of us would agree (or we should that there is a very small to non-existent relationship between gender and opinion.
   i. The percentage "for" is about the same.
5. Note also that we have only interviewed N = 80 people and that each column contains 40 respondents.
6. Note finally that the chi square is .200 with a reported P of .655. This means that the observed chi square (.200) **is** likely if the variables are statistically independent in the population.
7. So far so good.

Table 1

Opinion/ Gender	Male	Female	
For	47.5% 190	52.5% 21	
Against	52.5 21	47.5 19	
Totals	100 (**40**)	100 (**40**)	N = **80**
Chi square = .200 (P = .655)			

8. Now suppose we managed to interview 10 times as many people thereby having an N = 800 subjects.

Table 2			
Opinion/ Gender	Male	Female	
For	47.5% 190	52.5% 210	
Against	52.5 210	47.5 190	
Totals	100 (**400**)	100 (**400**)	N = **800**
Chi square = .2.00 (P = .157)			

i. The relationship as measured by the percentages has not changed at all. About 50 percent of both men and women are "for" the policy.
ii. So we would conclude that the substantive relationship remains very weak. Adding respondents hasn't changed the substantive of theoretical interpretation.
iii. But notice that the chi square has increased by a factor of 10.
iv. And the probability of observing this chi square has decreased. It's still above our .05 level or cut off point. Nevertheless, it's seems to suggest that the null hypothesis was close to being rejected.

9. Now suppose we had unlimited riches and could interview without end.
   i. Suppose in fact we interviewed N = 8,000 individuals but that the relationship between gender and opinion did not change.
   ii. The new table might be:

Table 3			
Opinion/ Gender	Male	Female	
For	47.5% 1900	52.5% 2100	
Against	52.5 2100	47.5 1900	
Totals	100 (**4000**)	100 (**4000**)	N = **8000**
Chi square = .20.00 (P = .000)			

iii. The sample size has been increased once again by 10 so the chi square increases by a factor of 10.
1) It's increased from 2.0 to 20.0.
2) **This will always be the case.** As N increases, chi square increases in lockstep, assuming all else is equal.

iv. So, now we still have weak relationship as measured by the percentages but the chi square is highly (statistically) significant.

F. Summary:
1. In the social sciences we have to separate two ideas:
i. **Statistical significance**: can a statistical hypothesis be rejected? If so, the result is **statistically significant**.
1) We use tests of significance like chi square to answer that question.
ii. Are the results meaningful from a policy or applied or commonsense or theoretical point of view?
1) We use percentages, magnitudes of differences, and measures of association to answer this question.

V. TEST YOURSELF:
A. Here are a table that shows a people's **beliefs** about pornography over time.
1. The respondents were asked "Does pornography lead to rape?" They could of course answer "yes" or "no"

Cells contain: -Column percent -N of cases		year								
		76	78	80	84	86	88	90	94	*ROW TOTAL*
pornrape	1 YES	**58.0** 787	**61.2** 867	**59.5** 790	**59.8** 806	**61.0** 830	**61.4** 558	**63.1** 507	**53.5** 491	***59.7*** *5,636*
	2 NO	**42.0** 569	**38.8** 549	**40.5** 538	**40.2** 542	**39.0** 530	**38.6** 351	**36.9** 297	**46.5** 426	***40.3*** *3,802*
	COL TOTAL	***100.0*** *1,356*	***100.0*** *1,416*	***100.0*** *1,328*	***100.0*** *1,348*	***100.0*** *1,360*	***100.0*** *909*	***100.0*** *804*	***100.0*** *917*	***100.0*** *9,438*
Means		1.42	1.39	1.41	1.40	1.39	1.39	1.37	1.46	1.40
Std Devs		.49	.49	.49	.49	.49	.49	.48	.50	.49

2. The chi square for this table is 23.25 (p= 0.00).

B. Questions:
   1. Has public opinion changed over time? Explain.
   2. (For 3 points of extra credit) Is there a substantively meaningful change as opposed to a statistically significant one? Explain.
      i. You need to study the material in the preceding section of the notes.

VI. NEXT TIME:
   A. Discussion of research projects: political culture and change.
   B. Desktop statistics and presentations.
   C. Measures of association.

Measuring Social Class
First Try
Recodes

Old Codes:

1. Executives
2. Professionals
3. Technicians
4. Sales
5. Clerical
6. Private Household
7. Protective service
8. Service workers
9. Farmers
10. Precision (skilled) workers
11. Operators
12. Transportation
13. Laborers/helpers

14 Armed Forces

New Codes:

1. High status = 1 - 3
2. Modern proletariat = 4 - 5, 7
3. Traditional proletariat =7, 10-12
4. Low status = 6, 8, 13

Excluded: 14 armed forces