DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS
Research Methods Posc 302

MEASURES OF ASSOCIATION

I. TODAY'S SESSION:
- A. Discussion of papers
- B. Measures of association and correlation
- C. Writing tips:
    - 1. Look over previous notes: MAKE SURE THAT YOU DON'T MAKE ANY OBVIOUS ERRORS DISCUSSED UNDER "WRITING TIPS" IF YOU WANT A DECENT GRADE.

II. PAPERS:
- A. Organization:
    - 1. Extended introduction
    - 2. "Methods" section
        - i. Hypotheses
        - ii. Operational definitions
            - 1) Choice of variables
                - a) Point reader to appendix if necessary
        - iii. Data source and why it was chosen
        - iv. Choice of statistical methods.
            - 1) Make this part brief.
            - 2) Do not explain statistics; rather simply indicate which will be used.
    - 3. Analysis:
        - i. Assume that you are making a presentation.
        - ii. In the body of the paper make your case.
        - iii. Properly labeled tables
            - 1) The chi square statistic should always be reported with the degrees of freedom that will always be a part of any computer report.
            - 2) Report measures of association.
    - 4. Conclusion
        - i. Summarize what you found
        - ii. Discuss its importance
        - iii. What questions remain
            - 1)
            - 2) No three dimensional bars, please.
            - 3) No pie charts
        - iv. Scales and axes must be labeled.

III. CORRELATION COEFFICIENT - SUMMARY:

A. The correlation coefficient summarizes the strength and direction (positive or negative) of a correlation between X and Y, which are **numeric** or **quantitative** variables.

B. Its value lies between -1.0 and +1.0.

C. Interpretation:

1. 1.0 means perfect positive linear correlation.
   i. All data points in a scatterplot lie on a straight line that has a positive slope.
      1) That is, a line that slopes upward from lower left to upper right on the X-Y coordinate system.
2. r = .9 means a very strong positive linear correlation.
   i. Most point lie on a straight line that has a positive slope.
3. r = .75 means a strong positive correlation
4. r = .37 means a moderate to weak positive correlation
5. r = .15 means a weak correlation.
   i. Most point are scattered around a straight line, many of them being quite far from the line.
6. r = 0 means no linear correlation.
   i. **Important: r = 0 means X and Y are not linearly correlated. But they may have some sort of statistical dependence. r =0 does not mean that the variables are independent. It only means that they are not linearly correlated.**
   ii. Example:

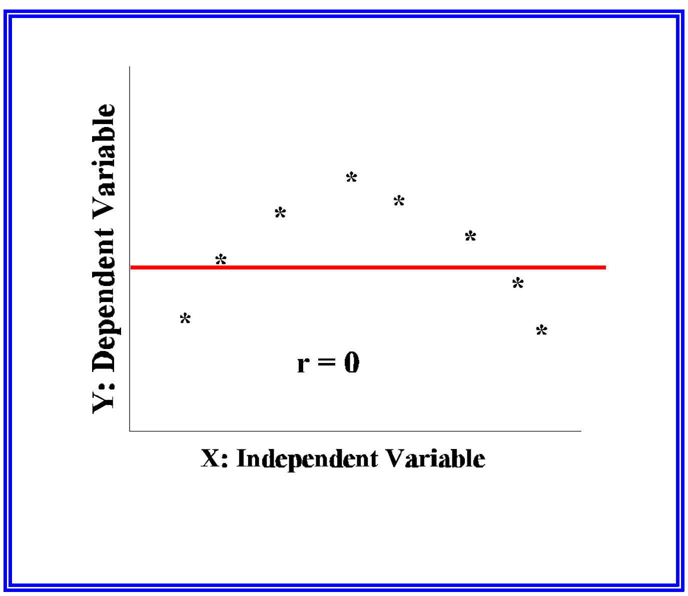

Figure 1: Relationship But No Correlation

iii. The figure (Figure 1) shows that if one knows the value of X, one can predict exactly the value of Y. So the variables are related.

iv. But there is not a **linear** correlation between them.

1) Just apply the definition: the greater the X, the greater the Y. It doesn't hold in this case, does it?

7. Negative values are interpreted in the same way except that the "direction" of the correlation is the opposite.

i. Example: r = -.27 means a weak negative correlation: as X increases, Y **tends** to decrease.

D. Note that strictly speaking r, the correlation coefficient, applies to quantitative variables.

1. But most if not all of our data involve at most ordinal variables.

IV. PROBLEM:

A. Here is a crosstabulation between party identification and attitudes toward women's role in society.

-Column percent -N of cases		1 Demo	2 Ind	3 Rep	*ROW* *TOTAL*
v960543	1 Women and men should have equal roles	**57.5** 356	**58.6** 220	**35.1** 153	***50.9*** *729*
	2 2	**17.8** 110	**16.8** 63	**19.7** 86	***18.1*** *259*
	3 3	**6.2** 38	**6.3** 24	**14.3** 62	***8.7*** *124*
	4 4	**9.9** 61	**9.0** 34	**15.0** 65	***11.2*** *161*
	5 5	**3.1** 19	**2.4** 9	**7.4** 32	***4.2*** *60*
	6 6	**1.9** 12	**4.1** 15	**4.4** 19	***3.2*** *46*
	7 A woman's place is in the home	**3.7** 23	**2.9** 11	**4.1** 18	***3.6*** *52*
	COL TOTAL	***100.0*** *619*	***100.0*** *376*	***100.0*** *437*	***100.0*** *1,432*

Figure 2: Relationship Between Party Identification and Attitudes Toward Women's Rights

1. We can use the percentages to help understand the nature of the relationship.

2. But it would be helpful to have an overall measure, a single number, that would tell us a lot of about how party identification and attitudes toward women's roles were related.

V. ORDINAL VARIABLES:

A. Although r does not apply to categorical data, we can use measures of correlation that have roughly the same interpretation.

1. We'll confine ourselves to strictly ordinal variables.

B. Ordinal variables: values are categories but the categories have an implicit or even explicit order.

1. Example 1:

i. Suppose we have three age groups: "Young" (less than 30 years); "Mature" (30 to 55 years); and "Old Coot" (over 56 years).

1) Although the values of this variable are categories, they can be "ordered" from lowest to highest.

2) Hence, the variable is ordinal.

2. Example 2:

i. Consider party identification. The categories seem to be just names. But we could think of them as forming an implicit scale running from most Democratic to least Democratic.

Category	Amount of "Democratic-ness"
Strong Democrat	Most (a lot)
Democrat	Some
Independent	Very little, if any
Republican	None
Strong Republican	None at all

1) The long and short is that we can argue that party identification is an ordinal variable.

ii. Example 3:

1) Scale of support for women's rights runs from 1 ("Women and men should have equal roles") to 7 ("A woman's place is in the home").

2) This can be thought of as an ordinal variable since there is an implicit order of magnitude running from a lot of support for women's rights to none.

C. ***<u>Note that a "dichotomy"–a variable with just two categories–can always be considered ordinal.</u>***

VI. MEASURES OF ORDINAL CORRELATION:

A. Without going into details we can assert that there are several different ordinal measures of correlation that are similar to the correlation coefficient.

1. They are called:

i. Gamma

ii. Tab-b and tau-c

B. Strictly speaking each has its own particular meaning since each is defined mathematically in a different way.

C. Nevertheless, their numerical interpretation follows the ideas outlined above and in class 21.

1. They are bounded: their values lie between -1.0 and 1.0.
2. The close they are to 1.0, the stronger the correlation and hence relationship between X and Y.
3. Values near 0 suggest a very weak correlation or even no correlation.

D. Look at the sign of the coefficients and interpret as with r.

E. It's easier to understand how to use these measures by looking at a couple of concrete examples.

1. Party identification by vote

Cells contain: -Column percent -N of cases		v960420			
		1 Demo	2 Ind	3 Rep	*ROW TOTAL*
v961082	1 Bill Clinton	**94.7** 385	**79.2** 111	**12.8** 40	***62.3*** *536*
	2 Bob Dole	**5.3** 21	**20.8** 29	**87.2** 274	***37.7*** *324*
	COL TOTAL	***100.0*** *406*	***100.0*** *140*	***100.0*** *314*	***100.0*** *860*
	Means	1.05	1.21	1.87	1.38
	Std Devs	.22	.41	.33	.48

Figure 3: Party Identification By Vote

i. The interpretation of the table can be aided by looking at the measures of association that accompany the SDA report.

Summary Statistics					
Eta* =	.78	Gamma =	.94	Chisq(P) =	569.06 (p= 0.00)
R =	.76	Tau-b =	.72	Chisq(LR) =	636.51 (p= 0.00)
Somers' d* =	.62	Tau-c =	.77	df =	2

*Column variable treated as the dependent variable.

Figure 4: Measures of Ordinal Correlation

ii. Note that gamma is about 1.0. The value is .94, which is close to

1.0.

1) This suggests a strong correlation between vote and party identification.

2) In this instance the party identification variable is interpreted as "strength of Republican" identification and hence gamma = .94 means that the more Republican a person is, the more that person "votes" for Dole.

a) Look at the table.

iii. Similarly, tau-b and tau-c are "large."

1) **In the context of ordinal data analysis a value of .5 or above suggests a moderate to strong correlation.**

2) Tau-b equals .72 and tau-c is .77 so there is again evidence of a strong positive correlation between the two variables.

2. Party identification by opinion about homosexuals serving in the military.

Frequency Distribution						
Cells contain: -Column percent -N of cases		v960420			
		1 Demo	2 Ind	3 Rep	*ROW TOTAL*
v961195	1 Homosexuals should be allowed to serve	**79.0** 442	**75.2** 240	**50.9** 200	***69.3*** *883*
	5 Homosexuals should not be allowed to serve	**21.0** 117	**24.8** 79	**49.1** 193	***30.7*** *390*
	COL TOTAL	***100.0*** *559*	***100.0*** *320*	***100.0*** *394*	***100.0*** *1,273*
	Means	1.84	1.99	2.96	2.23
	Std Devs	1.63	1.73	2.00	1.84

Figure 5: Partisanship By Attitudes Towards Gays in the Military

i. It looks like there is a relationship, but how strong is it? Here are the summary measures:

Summary Statistics						
Eta* =	.27	Gamma =	.42	Chisq(P) =	95.81	(p= 0.00)
R =	.25	Tau-b =	.23	Chisq(LR) =	92.89	(p= 0.00)
Somers' d* =	.19	Tau-c =	.25	df =	2	

*Column variable treated as the dependent variable.

Figure 6: Measures of Ordinal Correlation

1) Gamma, tau-b, and tau-c coefficients are relatively “modest,” which suggests a weak to moderate relationship.
2) Party is related to attitudes on this issues, but there isn’t a very strong connection.
   a) It appears that this “social issue” doesn’t divide party followers as much as perhaps party elites.

3. Partisanship by watching “ER,” the popular television.
   i. Off hand I can think of a reason why Democrats would be more or less likely to watch ER that Republicans or independents, but you never know.
   ii. The cross-classification is

Cells contain: -Column percent -N of cases		v960420			
		1 Demo	2 Ind	3 Rep	*ROW TOTAL*
v961150	1 Every week	**11.3** 66	**7.6** 25	**13.9** 56	***11.2*** *147*
	2 Most weeks	**8.2** 47	**8.3** 27	**8.9** 36	***8.4*** *111*
	3 Only occasionally	**27.1** 158	**29.5** 98	**23.6** 96	***26.7*** *351*
	4 Not at all	**53.4** 310	**54.6** 180	**53.5** 216	***53.7*** *706*
	COL TOTAL	***100.0*** *581*	***100.0*** *330*	***100.0*** *404*	***100.0*** *1,315*

Figure 7: Party Identification By ER Viewership

1) Looks like no relationship.
2) The chi square and summary measures confirm this impression:

Summary Statistics						
Eta* =	.05	Gamma =	-.01	Chisq(P) =	9.53	(p= 0.15)
R =	-.02	Tau-b =	-.01	Chisq(LR) =	9.86	(p= 0.13)
Somers' d* =	-.01	Tau-c =	-.01	df =	6	

*Column variable treated as the dependent variable.

Figure 8: Ordinal Summary Measures

3) Note that gamma is nearly 0, as are tau-b and tau-c. These figures suggest no correlation between the two variables, as

we would expect

VII. PRESENTATION:

A. Here once again is what your tables should look like. You don't have to follow this pattern exactly. But if any elements are missing, your grade will suffer.

Table 4
Party Identification By Region
(Percentages)

	North	Midwest	West
Democrat	**58**	**44**	**38**
Republican	**42**	**56**	**62**
Total	**100 (431)**	**100 (337)**	**100 (401)**

For question wording see Appendix B.
Data: General Social Survey Cumulative File

VIII. NEXT TIME:

A. Final wrap up.

B. Reading:

1. Johnson and Joslyn, ***Research Methods***, pages 336 to 367.
   i. Make an effort to understand the meaning and use of the measures of association.