

Lecture 14:

Elaborating bivariate tables

Ernesto F. L. Amaral

Advanced Methods of Social Research (SOCI 420)

www.ernestoamaral.com

Source: Healey, Joseph F. 2015. "Statistics: A Tool for Social Research." Stamford: Cengage Learning. 10th edition. Chapter 14 (pp. 380–404).



Chapter learning objectives

- Explain the purpose of multivariate analysis in terms of observing the effect of a control variable
- Construct and interpret partial tables
- Compute and interpret partial measures of association
- Recognize and interpret direct, spurious or intervening, and interactive relationships
- Compute and interpret partial gamma
- Explain limitations of elaborating bivariate tables



Controlling for a third variable

- Social science research projects are multivariate
- One way to conduct multivariate analysis is to observe the effect of third variables, one at a time, on a bivariate correlation
- The elaboration technique extends the analysis of bivariate tables and associations

Partial tables

- We observe how a control variable (Z) affects the relationship between X and Y
- To control for a third variable, the bivariate relationship is reconstructed for each value of the control variable
- Tables that display the relationship between X and Y for each value of Z (a third variable) are called partial tables

Focus on three basic patterns

- Direct relationships
- Spurious or intervening relationships
- Interaction



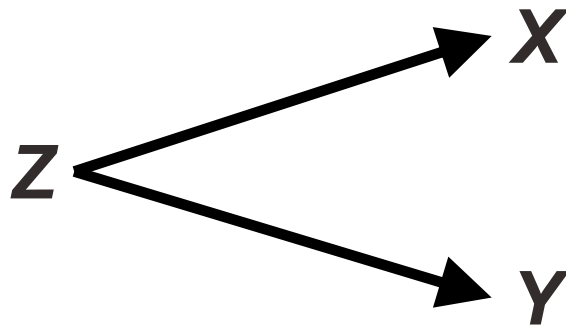
Direct relationships

- In a direct relationship, the control variable has little effect on the relationship between X and Y
- The column percentages and Gammas in the partial tables are about the same as the bivariate table
- This outcome supports the argument that X causes Y
- Also referred to as replication



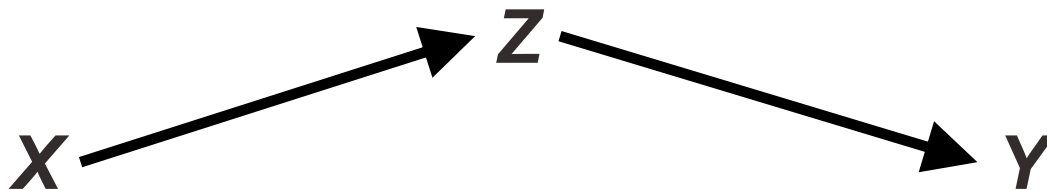
Spurious relationships

- In a spurious relationship, X and Y are not related, both are caused by Z
- In a spurious relationship, the Gammas in the partial tables are dramatically lower than the gamma in the bivariate table, perhaps even falling to zero
- Also referred to as explanation



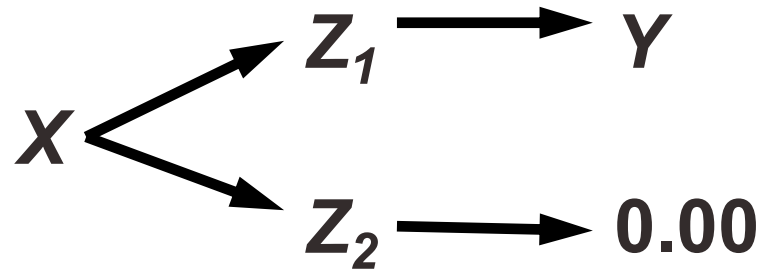
Intervening relationships

- In an intervening relationship, X and Y are not directly related to each other but are linked by Z , which “intervenes” between the two
- Also referred to as interpretation

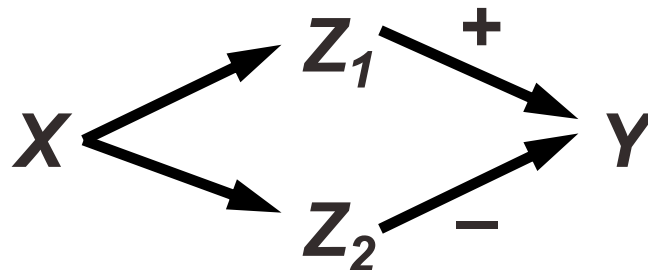


Interaction

- Interaction occurs when the association between X and Y changes across the categories of Z
 - X and Y could only be related for some categories of Z



- X and Y could have a positive association for one category of Z and a negative association for others



Summary

- Possible results when controlling for third variables

Compared with Bivariate Table, Partial Tables Show	Pattern	Implications for Further Analysis	Likely Next Step	Theoretical Implications
Same relationship between X and Y (e.g., gammas for partial tables are no more than ± 0.10 different from the bivariate gamma)	Direct relationship (replication)	Disregard Z	Analyze another Z variable	Theory that X causes Y is supported
Weaker relationship between X and Y (e.g., gammas for partial tables are all at least 0.10 weaker than the bivariate gamma)	Spurious relationship (explanation)	Incorporate Z	Focus on relationship between Z and Y	Theory that X causes Y is not supported
	Intervening relationship (interpretation)	Incorporate Z	Focus on relationships among X, Y, and Z	Theory that X causes Y is partially supported but must be revised to take Z into account
Mixed (e.g., there is a difference of at least ± 0.10 between gammas for the partial tables and between the gammas for partial tables and the bivariate gamma)	Interaction (specification)	Incorporate Z	Analyze subgroups (categories of Z) separately	Theory that X causes Y is partially supported but must be revised to take Z into account

Partial Gamma

- Partial Gamma indicates the overall strength of association between X and Y after the effects of the control variable (Z) have been removed
 - Compare Partial Gamma (G_p) to the Gamma (G) for the bivariate table to see if the relationship has changed

$$G_p = \frac{\sum N_s - \sum N_d}{\sum N_s + \sum N_d}$$

- N_s is the number of pairs of cases ranked in the **same order** across all partial tables
- N_d is the number of pairs of cases ranked in **different order** across all partial tables

Example 1

- Association between
 - Number of memberships in student organizations
 - X, independent variable
 - Satisfaction with college
 - Y, dependent variable

Satisfaction with College by Number of Memberships in Student Organizations

Satisfaction (Y)	Memberships (X)		TOTALS
	<i>None</i>	<i>At Least One</i>	
Low	57 (54.3%)	56 (33.9%)	113
High	48 (45.7%)	109 (66.1%)	157
TOTALS	105 (100.0%)	165 (100.0%)	270

Gamma = +0.40

Interpretation

- Comparing the conditional distributions of Y (the column percentages), we find a positive relationship
 - This direction is confirmed by the sign of Gamma (+0.40)
- College students with at least one membership in a student organization are more likely than students with no memberships to have high satisfaction with college



GPA as a control variable

- Associations remain positive

Satisfaction by Membership, Controlling for GPA

A. High GPA			
Satisfaction (Y)	Memberships (X)		TOTALS
	<i>None</i>	<i>At Least One</i>	
Low	29 (54.7%)	28 (34.1%)	57
High	24 (45.3%)	54 (65.9%)	78
TOTALS	53 (100.0%)	82 (100.0%)	135

Gamma = 0.40

B. Low GPA			
Satisfaction (Y)	Memberships (X)		TOTALS
	<i>None</i>	<i>At Least One</i>	
Low	28 (53.8%)	28 (33.7%)	56
High	24 (46.2%)	55 (66.3%)	79
TOTALS	52 (100.0%)	83 (100.0%)	135


Gamma = 0.39



Association still positive

- The relationship between integration and satisfaction is the same in the partial tables as it was in the bivariate table
 - This is evidence of a **direct relationship**

High GPA	Low GPA
$N_s = (29)(54) = 1566$	$N_s = (28)(55) = 1540$
$N_d = (28)(24) = 672$	$N_d = (28)(24) = 672$

$$G_p = \frac{\sum N_s - \sum N_d}{\sum N_s + \sum N_d} = \frac{(1566 + 1540) - (672 + 672)}{(1566 + 1540) + (672 + 672)} = 0.40$$


Class standing as a control

- There is no more association
 - Upperclass students: seniors and juniors
 - Underclass students: sophomores and freshmen

Satisfaction by Membership, Controlling for Class

A. Upperclass Students			
Satisfaction (Y)	Memberships (X)		TOTALS
	None	At Least One	
Low	8 (25.0%)	32 (24.8%)	40
High	24 (75.0%)	97 (75.2%)	121
TOTALS	32 (100.0%)	129 (100.0%)	161

Gamma = 0.01

B. Underclass Students			
Satisfaction (Y)	Memberships (X)		TOTALS
	None	At Least One	
Low	49 (67.1%)	24 (66.7%)	73
High	24 (32.9%)	12 (33.3%)	36
TOTALS	73 (100.0%)	36 (100.0%)	109

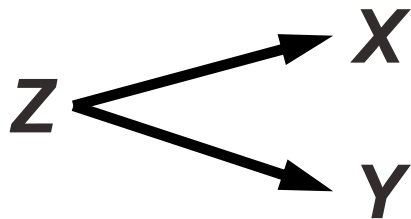
Gamma = 0.01



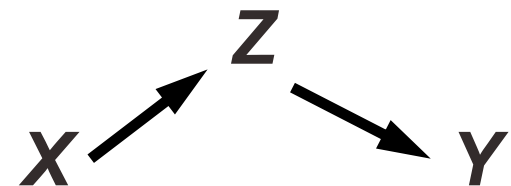
Association disappears

- The original bivariate relationship between memberships and satisfaction disappears in the partial tables
 - When the association disappears, we have either a spurious or an intervening relationship

Spurious



Intervening



Spurious relationship

- Decision about whether the association is spurious or intervening is based on
 - Temporal (timing) or theoretical grounds
- A spurious relationship makes more sense
 - Class standing likely predicts the number of memberships, and not the other way around
 - Partial Gamma supports our conclusion (reduced to zero)

Upperclass	Underclass
$N_s = (8)(97) = 776$	$N_s = (49)(12) = 588$
$N_d = (32)(24) = 768$	$N_d = (24)(24) = 576$

$$G_p = \frac{\sum N_s - \sum N_d}{\sum N_s + \sum N_d} = \frac{(776 + 588) - (768 + 576)}{(776 + 588) + (768 + 576)} = 0.01$$

Example 2

- Relationship for 50 immigrants between
 - Length of residence: X , independent variable
 - English fluency: Y , dependent variable

English Fluency	Length of Residence		TOTALS
	<i>Less Than Five Years (Low)</i>	<i>More Than Five Years (High)</i>	
Low	20	10	30
High	5	15	20
TOTALS	25	25	50

- $\text{Gamma} = +0.67$
 - Strong and positive association
 - As length of residence increases, English fluency also increases



Sex as a control variable

- Associations remain positive

- $G_m = 0.78$

A. Males			
	Length of Residence		
English Fluency	<i>Less Than Five Years (Low)</i>	<i>More Than Five Years (High)</i>	TOTALS
Low	10	5	15
High	2	8	10
TOTALS	12	13	25

- $G_f = 0.65$

B. Females			
	Length of Residence		
English Fluency	<i>Less Than Five Years (Low)</i>	<i>More Than Five Years (High)</i>	TOTALS
Low	10	5	15
High	3	7	10
TOTALS	13	12	25



Partial Gamma

$$G_m = \frac{N_s - N_d}{N_s + N_d} = \frac{80 - 10}{80 + 10} = 0.78$$

$$G_f = \frac{N_s - N_d}{N_s + N_d} = \frac{70 - 15}{70 + 15} = 0.65$$

$$G_p = \frac{\sum N_s - \sum N_d}{\sum N_s + \sum N_d} = \frac{(80 + 70) - (10 + 15)}{(80 + 70) + (10 + 15)} = 0.71$$



Sex has no effect

- While the two Gammas for the partial tables (0.78 and 0.65) differ slightly
 - They both indicate a strong and positive association between length of residence and English fluency
- Comparing Partial Gamma (0.71) to the original Gamma (0.67), we find little difference
- We have evidence of a **direct relationship**
 - Controlling for sex does not affect the association between length of residence and English fluency for immigrants



Example 3

- Relationship for 78 juvenile males between
 - Academic record: X , independent variable
 - Delinquency: Y , dependent variable

Delinquency by Academic Record

Delinquency	Academic Record		TOTALS
	<i>Poor</i>	<i>Good</i>	
Low	13 (27.1%)	20 (66.7%)	33 (42.3%)
High	35 (72.9%)	10 (33.3%)	45 (57.7%)
TOTALS	48 (100.0%)	30 (100.0%)	78 (100.0%)

Gamma = -0.69

- Gamma = -0.69
 - Juvenile males with better academic records have lower delinquency



Area of residence as a control

- Associations differ across partial tables

Delinquency by Academic Record, Controlling for Residence

A. Urban			
Delinquency	Academic Record		TOTALS
	Poor	Good	
Low	10 (27.8%)	3 (30.0%)	13 (28.3%)
High	26 (72.2%)	7 (70.0%)	33 (71.7%)
TOTALS	36 (100.0%)	10 (100.0%)	46 (100.0%)

Gamma = -0.05

B. Nonurban			
Delinquency	Academic Record		TOTALS
	Poor	Good	
Low	3 (25.0%)	17 (85.0%)	20 (62.5%)
High	9 (75.0%)	3 (15.0%)	12 (37.5%)
TOTALS	12 (100.0%)	20 (100.0%)	32 (100.0%)

Gamma = -0.89



Interpretation

- Gamma for urban areas is -0.05
 - No association between academic record and delinquency
- Gamma for nonurban areas is -0.89
 - Strong and negative association between academic record and delinquency
- Associations between X and Y differ across partial tables
 - This is an indication of **interaction**

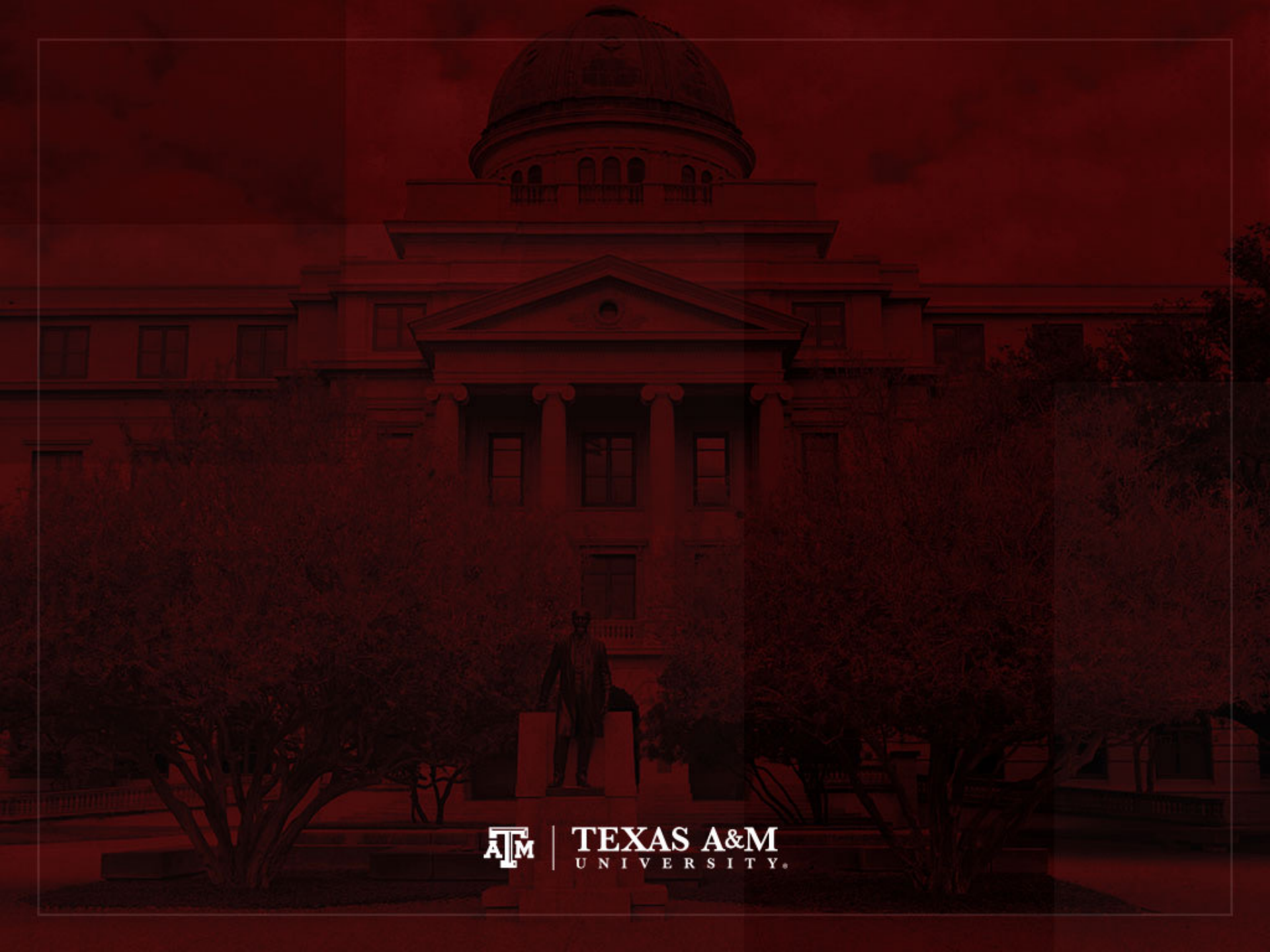


Origin of control variables

- Control variables are based on theory
- Research projects are anchored in theory, so control variables come mainly from theory
- Understanding a spurious relationship (explanation) or an intervening relationship (interpretation) cannot be based on statistical grounds or inspecting the partial tables

Limitations of partial tables

- Basic limitation: Sample size
 - Greater the number of partial tables, the more likely to run out of cells or have small cells
- Potential solutions
 - Reduce number of cells by collapsing categories (recoding)
 - Use very large samples
 - Use techniques appropriate for interval-ratio level



TEXAS A&M
UNIVERSITY.