

CONVENTIONAL UNDERSTANDING OF SETS

- Sets are binary, nominal-scale variables, the lowest and most primitive form of social measurement.
- The cross-tabulation of two sets is the simplest and most primitive form of variable-oriented analysis.
- This form of analysis is of limited value because: (1) the strength of the association between two binary variables is powerfully influenced by how they are created (e.g., the choice of cut-off values), and (2) with binary variables researchers can calculate only relatively simple measures of association. These coefficients may be useful descriptively, but they tell us little about the contours of relationships.
- In short, examining relations between binary variables might be considered adequate as a descriptive starting point, but this approach is too crude to be considered *real* social science.

DECONSTRUCTING THE 2X2 TABLE

CONVENTIONAL VIEW		
	Cause absent	Cause present
Outcome present	cases in this cell (#1) contribute to error	many cases should be in this cell (#2)
Outcome absent	many cases should be in this cell (#3)	cases in this cell (#4) contribute to error

- The conventional 2X2 table crosstabulating a causal condition and an outcome contains within it two important set-theoretic relations: the cause as a subset of the outcome and the outcome as a subset of the cause.
- These two kinds of set-theoretic relations involve *explicit* as opposed to *tendential* connections.
- The fact that there are two kinds of set-theoretic relations embodied in the conventional 2X2 table is completely outside the scope (and grasp) of conventional quantitative analysis. Conventional approaches focus on all four cells simultaneously.

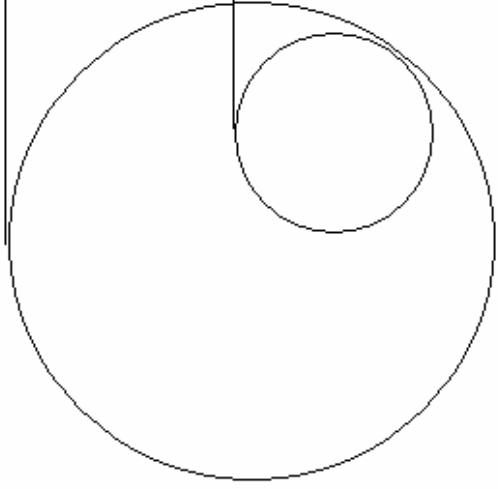
OUTCOME IS A SUBSET OF THE CAUSE		
	Cause absent	Cause present
Outcome present	1. no cases here	2. cases here
Outcome absent	3. cases here	4. cases here

Whenever the outcome is present the cause is present. The outcome is a subset of the cause.

CAUSE IS A SUBSET OF THE OUTCOME		
	Cause absent	Cause present
Outcome present	1. cases here	2. cases here
Outcome absent	3. cases here	4. no cases here

Whenever the cause is present the outcome is present. The cause is a subset of the outcome.

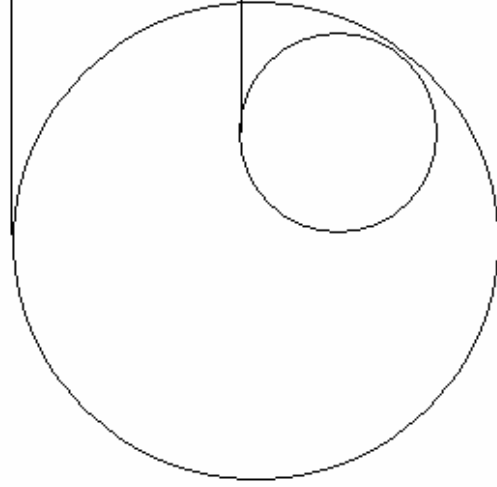
Political
Conservatives



The cause is a subset of the outcome.

Religious
Fundamentalists

People with Advanced
Degrees



The outcome is a subset of the cause.

Professionals

SETS AND SOCIAL SCIENTIFIC DISCOURSE

Many, if not most, social scientific statements, especially empirical generalizations about cross-case patterns, involve set-theoretic relationships:

- A. Religious fundamentalists are politically conservative. (Religious fundamentalists are a subset of politically conservative individuals.)
- B. Professionals have advanced degrees. (Professionals are a subset of those with advanced degrees.)
- C. Democracy requires a state with at least medium capacity. (Democratic states are a subset of states with at least medium capacity.)
- D. "Elite brokerage" is central to successful democratization. (Instances of successful democratization are a subset of instances of elite brokerage.)
- E. "Coercive" nation-building was not an option for "late-forming" states. (States practicing coercive nation-building are a subset of states that formed "early.")

Usually, but not always (e.g., D), the subset is mentioned first. Sometimes, it takes a little deciphering to figure out the set-theoretic relationship, as in E.

ELABORATING SET-THEORETIC ANALYSIS: THE SUPERSSET RELATION

- Conventional quantitative social science seeks additional variables to account for the error cases in cells 1 and 4. For example, a second binary variable that has cases with a value of 1 when $X = 0$ and $Y = 1$ and/or cases with a value of 0 when $X = 1$ and $Y = 0$ is a suitable second variable. The key is to use additional variables to achieve some sort of improved fit.
- In a superset analysis the goal is to move cases from cell 1 to cell 2 (i.e., to empty cell 1 of cases and thereby establish an explicit connection). In effect, the causal argument must be made more inclusive, which can be accomplished using logical *or*. Generally, this use of logical *or* entails moving up the ladder of abstraction to a more general conceptualization of the causal condition or construct.
- For example, to be hired one might need to demonstrate either training or experience. The set of individuals hired (the outcome) could then be viewed as a subset of those who have either (or both) of these characteristics.

ILLUSTRATION OF USE OF LOGICAL *OR* TO IDENTIFY EXPLICIT CONNECTION

	X Absent	X Present			both absent	X or Z Present
Outcome Present	5	25		Outcome Present	0	30
Outcome Absent	15	15		Outcome Absent	12	18

By identifying a substitutable causal condition (and moving to a more general conceptualization), it is possible to identify an explicit connection--the outcome is a subset of the reconstructed cause.

ELABORATING SET-THEORETIC ANALYSIS: THE SUBSET RELATION

- In a subset analysis, the goal is to move cases from cell 4 to cell 3 (i.e., to empty cell 4 of cases and thereby establish an explicit connection). In effect, the causal argument must be made more restrictive, which is accomplished through logical *and*. Generally, this use of logical *and* also entails moving down the ladder of abstraction to a more circumscribed conceptualization of the causal condition or construct.
- For example, to be fired from a position one might need to exhibit not only laziness but also dishonesty. The set of individuals combining these two traits could be seen as a subset of those who have been fired (the outcome).

	X Absent	X Present			X or Z absent	X*Z Present
Outcome Present	16	14		Outcome Present	18	12
Outcome Absent	24	6		Outcome Absent	30	0

NECESSITY AND SUFFICIENCY AS SUBSET RELATIONS

Anyone interested in demonstrating necessity and/or sufficiency must address set-theoretic relations. Necessity and sufficiency cannot be assessed using conventional quantitative methods.

CAUSE IS NECESSARY BUT NOT SUFFICIENT		
	Cause absent	Cause present
Outcome present	1. no cases here	2. cases here
Outcome absent	3. not relevant	4. not relevant

CAUSE IS SUFFICIENT BUT NOT NECESSARY		
	Cause absent	Cause present
Outcome present	1. not relevant	2. cases here
Outcome absent	3. not relevant	4. no cases here

SUFFICIENCY (WITHOUT NECESSITY)

I. Expressed as a simple truth table:

<u>Cause</u>		<u>Outcome</u>
1		1
0		1
0		0

II. Expressed as an inequality:

(values of the cause) \leq (value of the outcome)

III. Expressed as a research strategy: Find instances of the causal condition (i.e., select on the independent variable) and assess their agreement on the outcome (i.e., make sure that the outcome does not vary substantially across instances of the cause). This strategy is central to most forms of qualitative research.

NECESSITY (WITHOUT SUFFICIENCY)

I. Expressed as a simple truth table:

<u>Cause</u>		<u>Outcome</u>
1		1
1		0
0		0

II. Expressed as an inequality:

(values of the outcome) \leq (value of the cause)

III. Expressed as a research strategy: Find instances of the outcome (i.e., select on the dependent variable) and assess their agreement on the causal condition (i.e., make sure that the cause does not vary substantially across instances of the outcome).

CAUSAL COMPLEXITY

Another important benefit of set theoretic analysis is that it is much more compatible with the analysis of causal complexity than conventional techniques.

Example: a researcher studies production sites in a strike-prone industry and considers four possible causes of strikes:

technology = the introduction of new technology
wages = stagnant wages in times of high inflation
overtime = reduction in overtime hours
sourcing = outsourcing portions of production

Possible findings include:

- (1) technology \rightarrow strikes
- (2) technology*wages \rightarrow strikes
- (3) technology + wages \rightarrow strikes
- (4) technology*wages + overtime*sourcing \rightarrow strikes

In (1) technology is necessary and sufficient; in (2) technology is necessary but not sufficient; in (3) technology is sufficient but not necessary; in (4) technology is neither necessary nor sufficient. The fourth is the characteristic form of causal complexity: no cause is either necessary or sufficient.

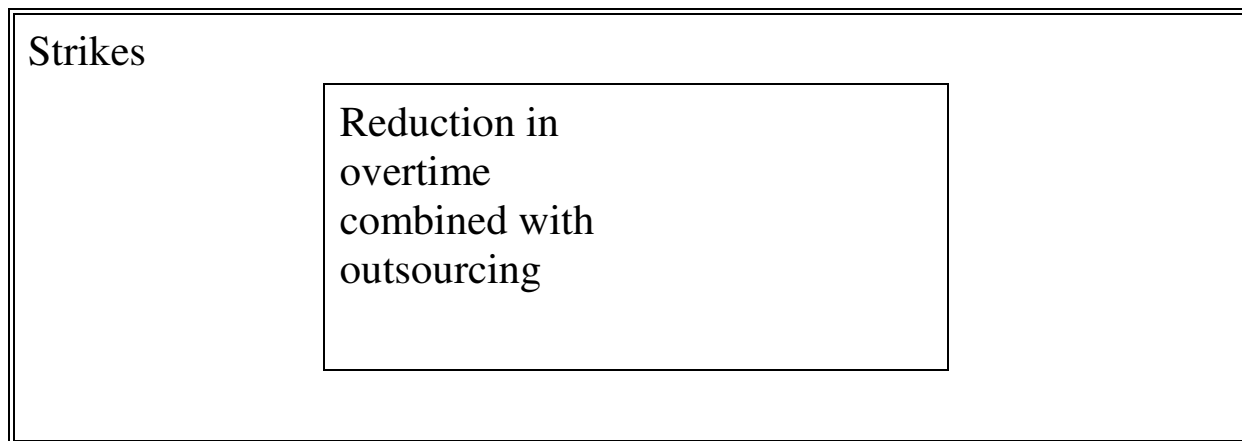
ASSESSING CAUSAL COMPLEXITY

I. Logical equation: technology*wages + overtime*sourcing → strikes

II. Formulated as a crosstabulation:

	Causal combination absent	Causal combination present
Strike present (1)	Cell 1: 20 cases	Cell 2: 23 cases
Strike absent (0)	Cell 3: 18 cases	Cell 4: 0 cases

III. Expressed as a Venn diagram:



The key to assessing the sufficiency of a combination of conditions, even if it is one among many combinations, is to select on instances of the combination and assess whether these instances agree on the outcome.

CASES AS CONFIGURATIONS

Set-theoretic methods also permit the examination of cases as configurations. This is usually accomplished through truth tables, which list the different logically possible combinations of conditions and the empirical evidence concerning each combination.

<i>C</i>	<i>L</i>	<i>H</i>	<i>G</i>	<i>U</i>	<i>N of Cases</i>
0	0	0	0	0	4
0	0	0	1	0	3
0	0	1	0	0	6
0	0	1	1	1	2
0	1	0	0	1	3
0	1	0	1	1	4
0	1	1	0	0	3
0	1	1	1	1	5
1	0	0	0	0	7
1	0	0	1	0	8
1	0	1	0	0	1
1	0	1	1	1	7
1	1	0	0	1	3
1	1	0	1	1	2
1	1	1	0	0	7
1	1	1	1	1	6

C = Corporatist wage negotiations

L = At least five years of rule by Left or Center-Left parties

H = Ethnic-cultural homogeneity

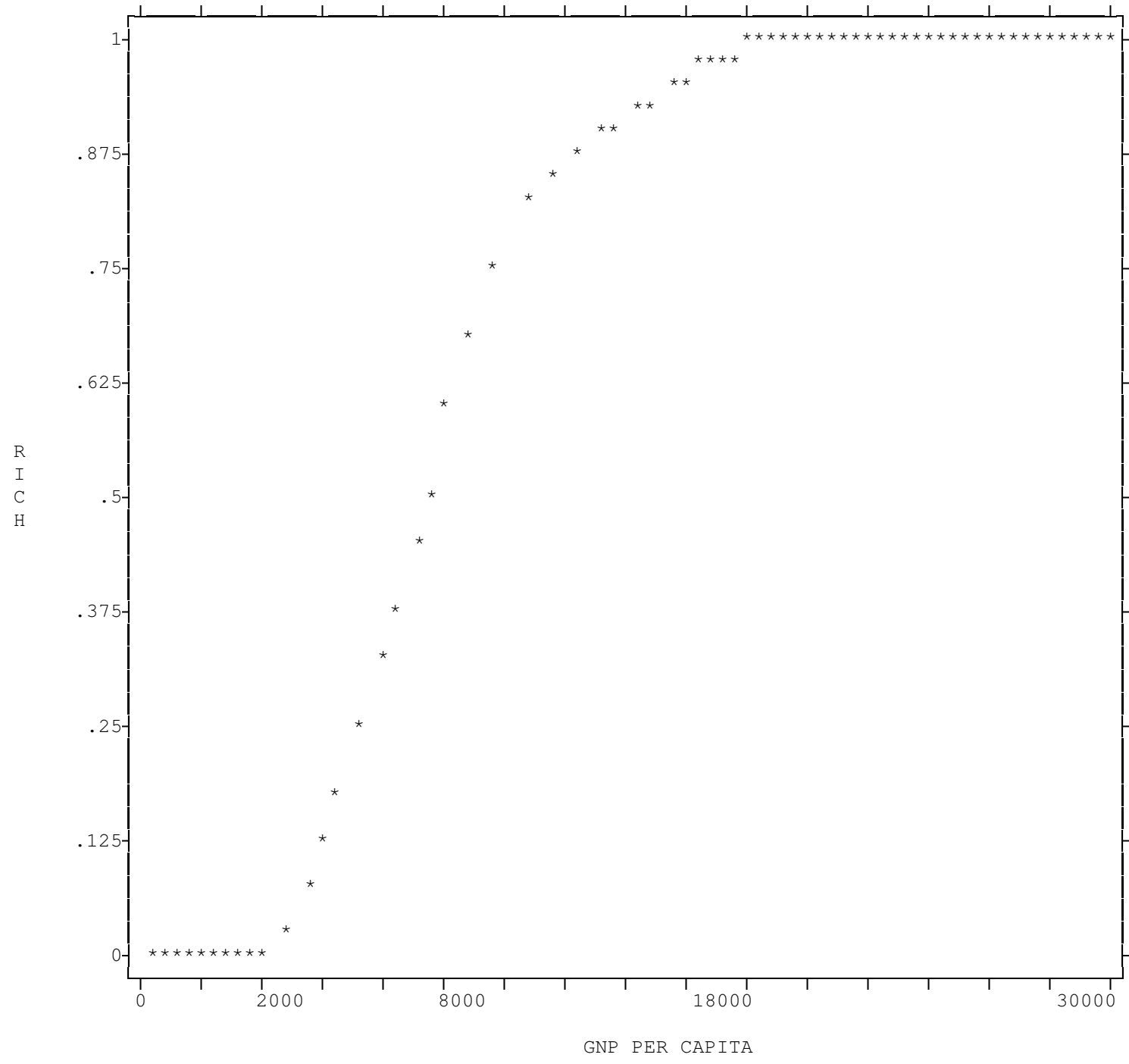
G = At least ten years of sustained economic growth

U = Adoption of universal pension system

FUZZY MEMBERSHIP IN THE SET OF "RICH COUNTRIES"

GNP/capita (US\$):	Membership (M):	Verbal Labels:
100 ----> 1,999	$M = 0$	clearly not-rich
2,000 ----> 7,999	$0 < M < .5$	more or less not-rich
8,000	$M = .5$	in between
8,001 ----> 17,999	$.5 < M < 1.0$	more or less rich
18,000 ----> 30,000	$M = 1.0$	clearly rich

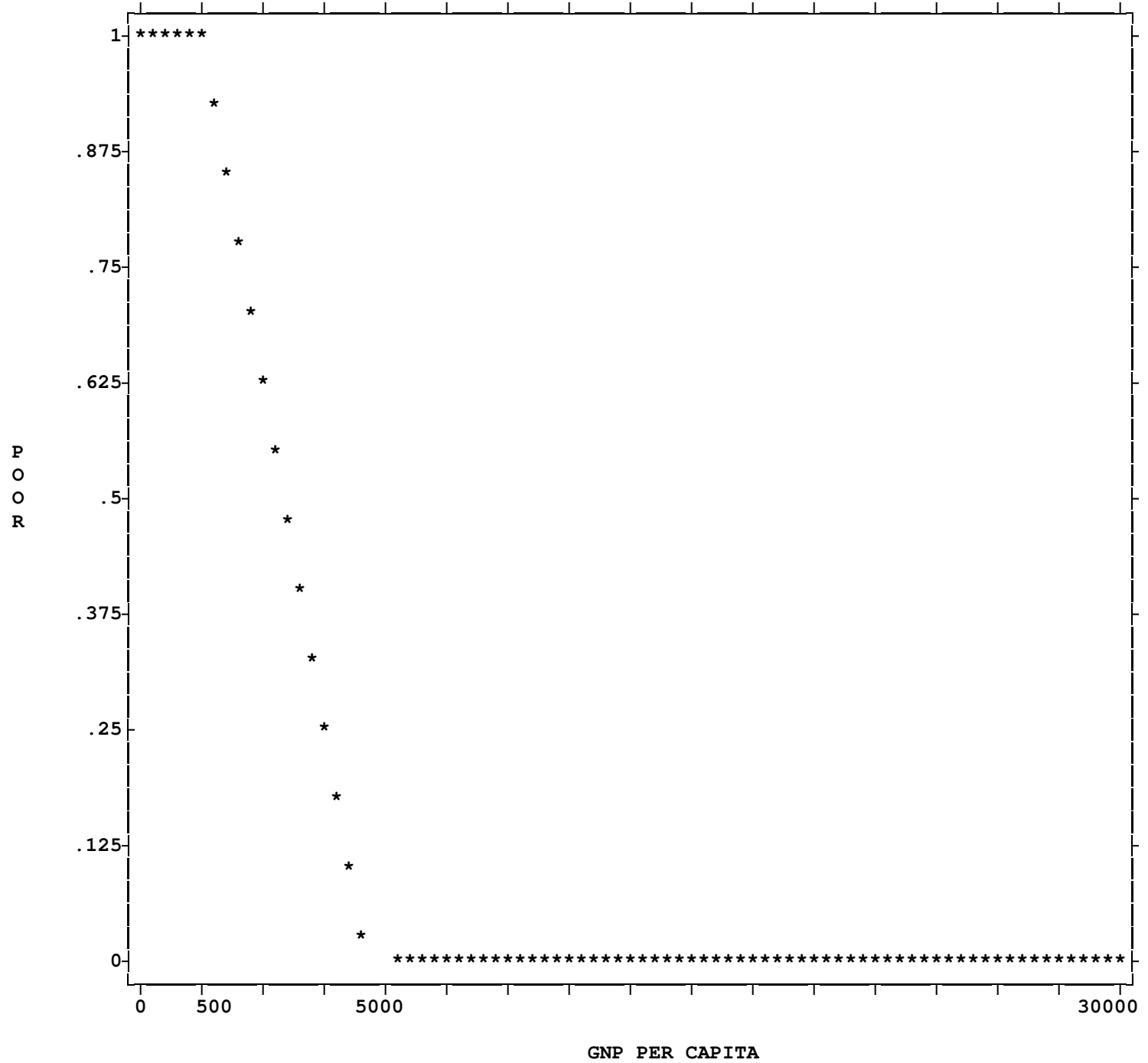
PLOT OF FUZZY MEMBERSHIP SCORES FOR THE SET OF "RICH COUNTRIES"



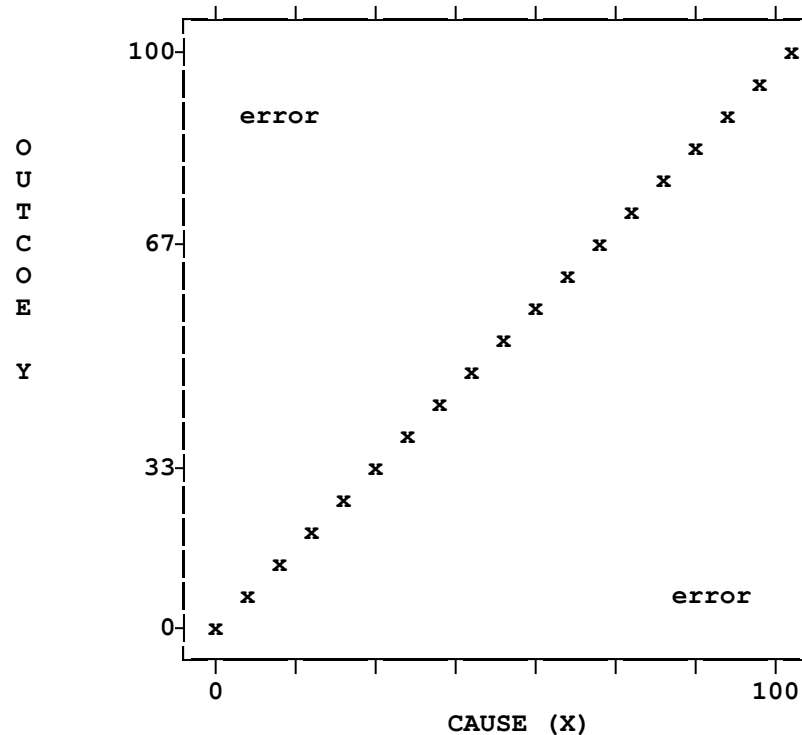
FUZZY MEMBERSHIP IN THE SET OF "POOR COUNTRIES"

GNP/capita (US\$):	Membership (M):	Verbal Labels:
100 ----> 499	$M = 1.0$	clearly poor
500 ----> 999	$.5 < M < .1$	more or less poor
1,000	$M = .5$	in between
1,001 ----> 4,999	$0 < M < .5$	more or less not-poor
5,000 ----> 30,000	$M = 0$	clearly not-poor

PLOT OF FUZZY MEMBERSHIP SCORES IN THE SET OF "POOR COUNTRIES"

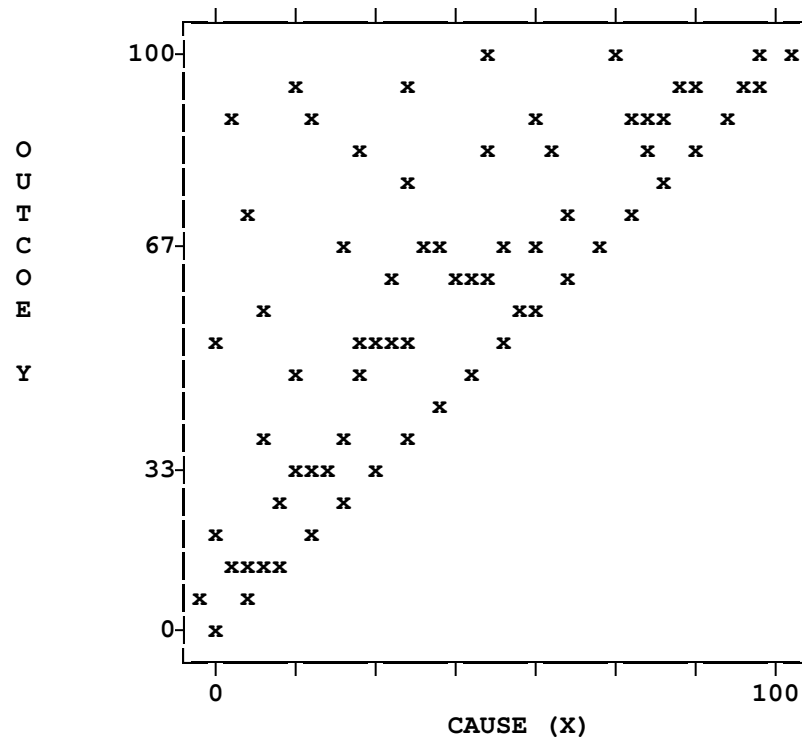


DECONSTRUCTING THE CONVENTIONAL SCATTERPLOT



- In conventional quantitative analysis, points in the lower-right corner and the upper-left corner of this plot are "errors," just as cases in cells 1 and 4 of the 2X2 crisp-set table were errors.
- With fuzzy sets, cases in these regions of the plot have different interpretations: Cases in the lower-right corner violate the argument that the cause is a subset of the outcome; cases in the upper-left corner violate the argument that the cause is a superset of the outcome (i.e., that the outcome is a subset of the cause).

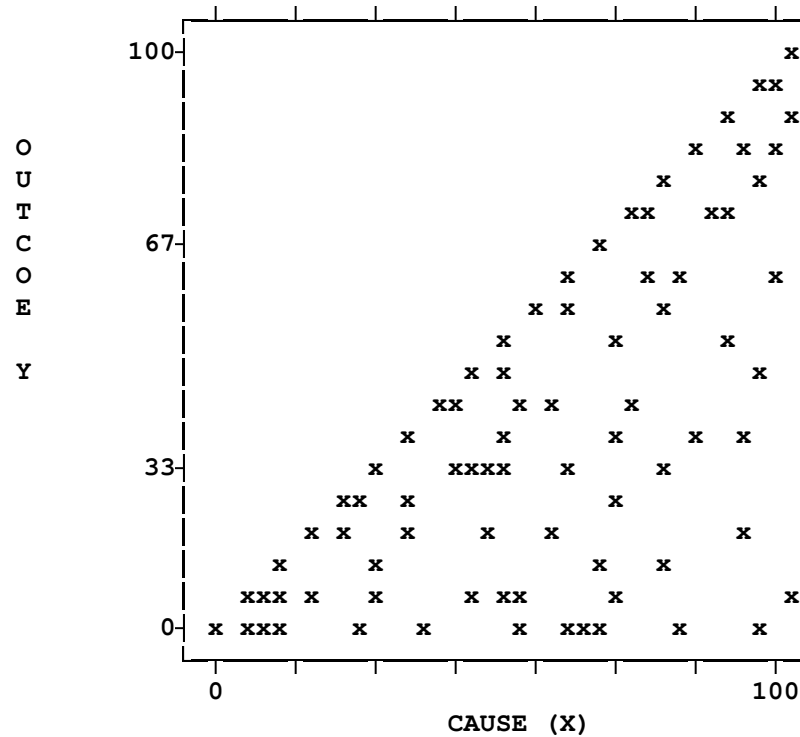
THE FUZZY SUBSET RELATION: THE CAUSE (X) IS A SUBSET OF THE OUTCOME (Y)



- This plot illustrates the characteristic upper-triangular plot indicating the fuzzy subset relation: $X \leq Y$ (cause is a subset of the outcome). This also can be viewed as a plot supporting the contention that X is sufficient for Y.

- Cases in the upper-left region are not errors, as they would be in a conventional quantitative analysis. Rather, these are cases with high membership in the outcome due to the operation of other causes. After all, the argument here is that X is a subset of Y (i.e., X is one of perhaps several ways to generate or achieve Y). Therefore, cases of Y without X (i.e., high membership in Y coupled with low membership in X) are to be expected.
- In this plot, cases in the lower-right region would be serious errors because these would be instances of high membership in the cause coupled with low membership in the outcome. Such cases would undermine the argument that there is an explicit connection between X and Y such that X is a subset of Y.

THE FUZZY SUPERSET RELATION: THE CAUSE (X) IS A SUPERSET OF THE OUTCOME (Y)



- This plot illustrates the characteristic lower-triangular plot indicating the fuzzy superset relation: $X \geq Y$ (cause is a superset of the outcome). This also can be viewed as a plot supporting the contention that X is necessary for Y.

- Cases in the lower-right region are not errors, as they would be in a conventional quantitative analysis. Rather, these are cases with low membership in the outcome, despite having high membership in the cause. This pattern indicates that Y is a subset of X : condition X must be present for Y to occur, but X may not be capable of generating Y by itself. Other conditions may be required as well. Therefore, cases of X without Y (i.e., high membership in X coupled with low membership in Y) are to be expected.
- Cases in the upper-left region would be serious errors because these would be instances of low membership in the cause coupled with high membership in the outcome. In this plot, such cases would undermine the argument that there is an explicit connection between X and Y such that X is a superset of Y (or Y is a subset of X).

ELABORATING THE FUZZY SUBSET RELATION

- When elaborating the subset relation with fuzzy sets, the goal is to move cases to the left side of the main diagonal of the scatterplot.
- When the argument is that the cause (X) is a subset of the outcome (Y), cases below the diagonal are "errors" because these X scores exceed the corresponding outcome (Y) scores.
- As with crisp set analysis, logical *and* can be used to move scores to the correct side of the diagonal. With logical *and*, conditions are compounded, which in turn involves taking the minimum membership score of the compounded sets as the membership of a case in the combinations. It follows mathematically that $A*B*C$ is less than or equal to $A*B$.
- Thus, the elaboration of a subset relation through additional compounded conditions lowers the X values and thus may move cases toward the left side of the diagonal.

FUZZY-SET DATA ON CLASS VOTING IN THE ADVANCED INDUSTRIAL SOCIETIES

Country	Weak Class voting (W)	Affluent (A)	Income Inequality (I)	Manufacturing (M)	Strong Unions (U)
Australia	0.6	0.8	0.6	0.4	0.6
Belgium	0.6	0.6	0.2	0.2	0.8
Denmark	0.2	0.6	0.4	0.2	0.8
France	0.8	0.6	0.8	0.2	0.2
Germany	0.6	0.6	0.8	0.4	0.4
Ireland	0.8	0.2	0.6	0.8	0.6
Italy	0.6	0.4	0.8	0.2	0.6
Netherlands	0.8	0.6	0.4	0.2	0.4
Norway	0.2	0.6	0.4	0.6	0.8
Sweden	0.0	0.8	0.4	0.8	1.0
United Kingdom	0.4	0.6	0.6	0.8	0.6
United States	1.0	1.0	0.8	0.4	0.2

ILLUSTRATION OF LOGICAL AND

Country	Affluent (A)	Income Inequality (I)		Manufacturing (M)		Strong Unions (U)	Affluent* Income Inequality	Affluent* Income Inequality* Weak Unions
	A	I	i	M	m	U	A*I	A*I*u
Australia	0.8	0.6	0.4	0.4	0.6	0.6	0.6	0.4
Belgium	0.6	0.2	0.8	0.2	0.8	0.8	0.2	0.2
Denmark	0.6	0.4	0.6	0.2	0.8	0.8	0.4	0.2
France	0.6	0.8	0.2	0.2	0.8	0.2	0.6	0.6
Germany	0.6	0.8	0.2	0.4	0.6	0.4	0.6	0.6
Ireland	0.2	0.6	0.4	0.8	0.2	0.6	0.2	0.2
Italy	0.4	0.8	0.2	0.2	0.8	0.6	0.4	0.4
Netherlands	0.6	0.4	0.6	0.2	0.8	0.4	0.4	0.4
Norway	0.6	0.4	0.6	0.6	0.4	0.8	0.4	0.2
Sweden	0.8	0.4	0.6	0.8	0.2	1.0	0.4	0.0
Ukingdom	0.6	0.6	0.4	0.8	0.2	0.6	0.6	0.4
Ustates	1.0	0.8	0.2	0.4	0.6	0.2	0.8	0.8

ELABORATING THE FUZZY SUPERSET RELATION

- When elaborating the superset relation with fuzzy sets, the goal is to move cases toward the right side of the main diagonal of the scatterplot.
- When the argument is that the cause (X) is a superset of the outcome (Y), cases above the diagonal are "errors" because these X scores are less than the corresponding outcome (Y) scores.
- As with crisp set analysis, logical *or* can be used to move scores to the correct side of the diagonal. With logical *or* conditions are substitutable, which in turn involves taking the maximum membership score of the substitutable sets. It follows mathematically that $A + B + C \geq A + B$.
- Thus, the elaboration of a superset relation through additional substitutable conditions raises the X values and thus moves cases toward the right side of the diagonal.

ILLUSTRATION OF LOGICAL *OR*

Country	Affluent (A)	Income Inequality (I)		Manufacturing (M)		Strong Unions (U)	Manufacturing + Strong Unions	Low Inequality * (Manufacturing + Strong Unions)
	A	I	i	M	m	U	M+U	i* (M+U)
Australia	0.8	0.6	0.4	0.4	0.6	0.6	0.6	0.4
Belgium	0.6	0.2	0.8	0.2	0.8	0.8	0.8	0.8
Denmark	0.6	0.4	0.6	0.2	0.8	0.8	0.8	0.6
France	0.6	0.8	0.2	0.2	0.8	0.2	0.2	0.2
Germany	0.6	0.8	0.2	0.4	0.6	0.4	0.4	0.2
Ireland	0.2	0.6	0.4	0.8	0.2	0.6	0.8	0.4
Italy	0.4	0.8	0.2	0.2	0.8	0.6	0.6	0.2
Netherlands	0.6	0.4	0.6	0.2	0.8	0.4	0.4	0.4
Norway	0.6	0.4	0.6	0.6	0.4	0.8	0.8	0.6
Sweden	0.8	0.4	0.6	0.8	0.2	1.0	1.0	0.6
UKingdom	0.6	0.6	0.4	0.8	0.2	0.6	0.8	0.4
UStates	1.0	0.8	0.2	0.4	0.6	0.2	0.4	0.2

FUZZY SETS AND CONFIGURATIONS

Country	Income Inequality		Manufacturing		Strong Unions									
	I	i	M	m	U	u	i^*m^*u	i^*m^*U	i^*M^*u	i^*M^*U	I^*m^*u	I^*m^*U	I^*M^*u	I^*M^*U
Australia	0.6	0.4	0.4	0.6	0.6	0.4	0.4	0.4	0.4	0.4	0.4	0.6	0.4	0.4
Belgium	0.2	0.8	0.2	0.8	0.8	0.2	0.2	0.8	0.2	0.2	0.2	0.2	0.2	0.2
Denmark	0.4	0.6	0.2	0.8	0.8	0.2	0.2	0.6	0.2	0.2	0.2	0.4	0.2	0.2
France	0.8	0.2	0.2	0.8	0.2	0.8	0.2	0.2	0.2	0.2	0.8	0.2	0.2	0.2
Germany	0.8	0.2	0.4	0.6	0.4	0.6	0.2	0.2	0.2	0.2	0.6	0.4	0.4	0.4
Ireland	0.6	0.4	0.8	0.2	0.6	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.4	0.6
Italy	0.8	0.2	0.2	0.8	0.6	0.4	0.2	0.2	0.2	0.2	0.4	0.6	0.2	0.2
Netherlands	0.4	0.6	0.2	0.8	0.4	0.6	0.6	0.4	0.2	0.2	0.4	0.4	0.2	0.2
Norway	0.4	0.6	0.6	0.4	0.8	0.2	0.2	0.4	0.2	0.6	0.2	0.4	0.2	0.4
Sweden	0.4	0.6	0.8	0.2	1.0	0.0	0.0	0.2	0.0	0.6	0.0	0.2	0.0	0.4
UKingdom	0.6	0.4	0.8	0.2	0.6	0.4	0.2	0.2	0.4	0.4	0.2	0.2	0.4	0.6
UStates	0.8	0.2	0.4	0.6	0.2	0.8	0.2	0.2	0.2	0.2	0.6	0.2	0.4	0.2

With three fuzzy sets, the vector space has eight corners. It is possible to calculate the membership of each case in each corner.

The corners can be viewed as ideal typic cases; the membership of a case in a corner is the degree to which it conforms to the ideal type represented by the corner.

In crisp-set analysis, by contrast, membership in a corner is either 1 or 0 and a case can have nonzero membership in only one corner.