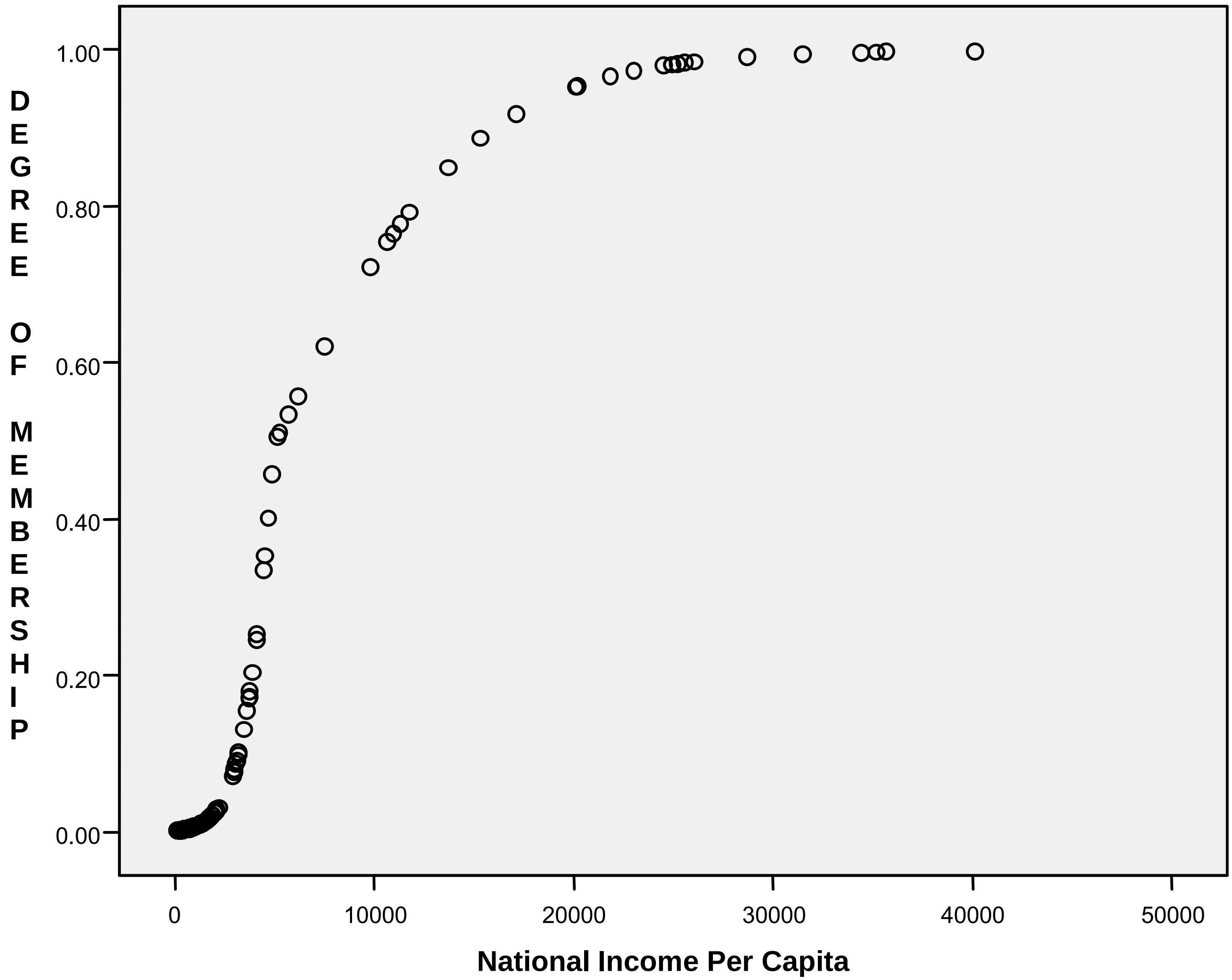
CRISP VERSUS FUZZY SETS

Crisp set	**Three-value fuzzy set**	**Four-value fuzzy set**	**Six-value fuzzy set**	**"Continuous" fuzzy set**
1 = fully in	1 = fully in	1 = fully in	1 = fully in	1 = fully in
			.8 = mostly but not fully in	Degree of membership is more "in" than "out": $.5 < x_i < 1$
		.75 = more in than out	.6 = more or less in	
	.5 = neither fully in nor fully out			.5 = cross-over: neither in nor out
		.25 = more out than in	.4 = more or less out	
			.2 = mostly but not fully out	Degree of membership is more "out" than "in": $0 < x_i < .5$
0 = fully out	0 = fully out	0 = fully out	0 = fully out	0 = fully out

FUZZY MEMBERSHIP IN THE SET OF "RICH" COUNTRIES

GNP/capita:	Membership (M):	Verbal Labels:
lowest $\rightarrow$ 2,499	$M = 0$	clearly not-rich
2,500 $\rightarrow$ 4,999	$0 < M < .5$	more or less not-rich
5,000	$M = .5$	cross-over point
5,001 $\rightarrow$ 19,999	$.5 < M < 1.0$	more or less rich
20,000 $\rightarrow$ highest	$M = 1.0$	clearly rich

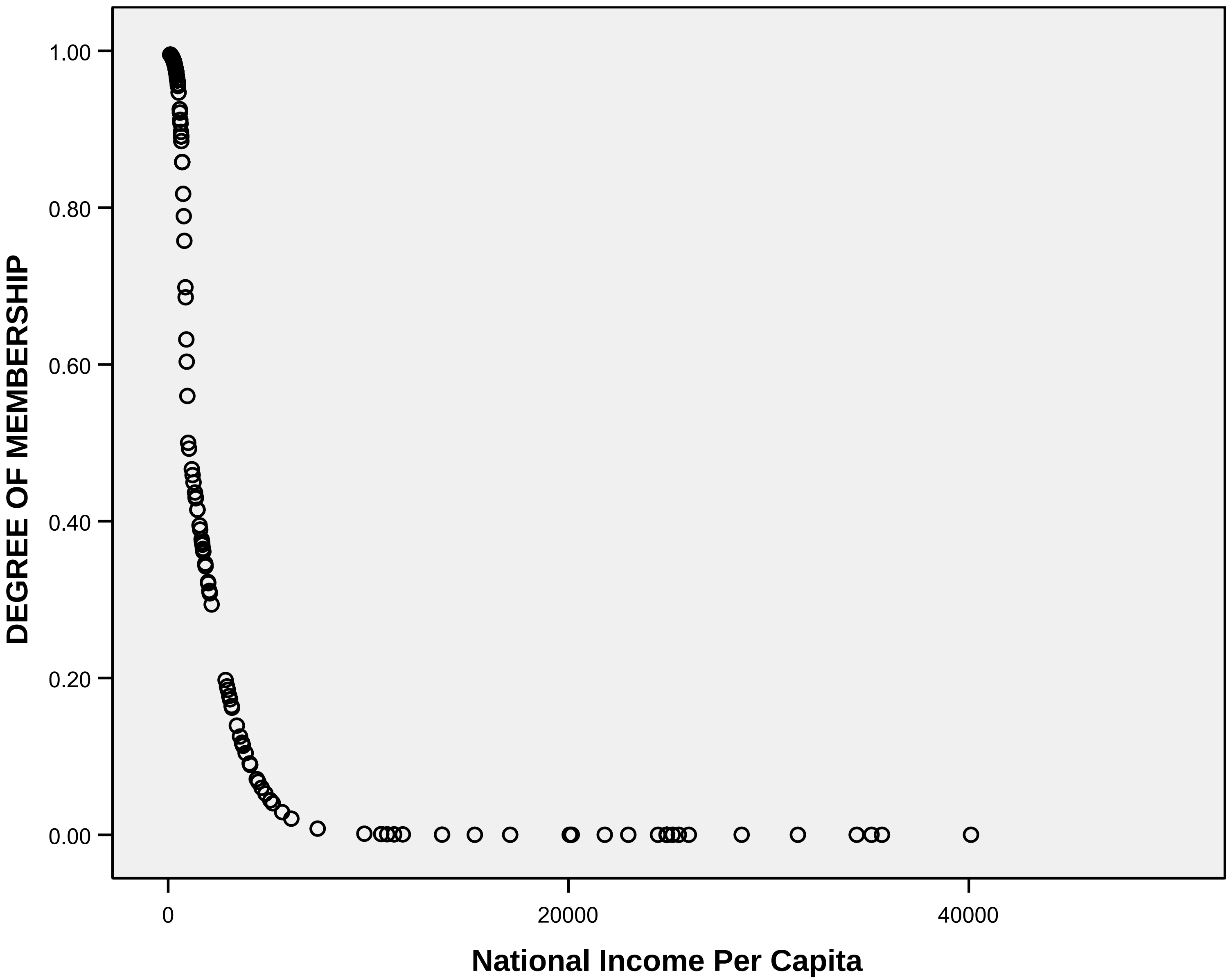
Plot of Degree of Membership in the Set of Rich Countries Against National Income Per Capita



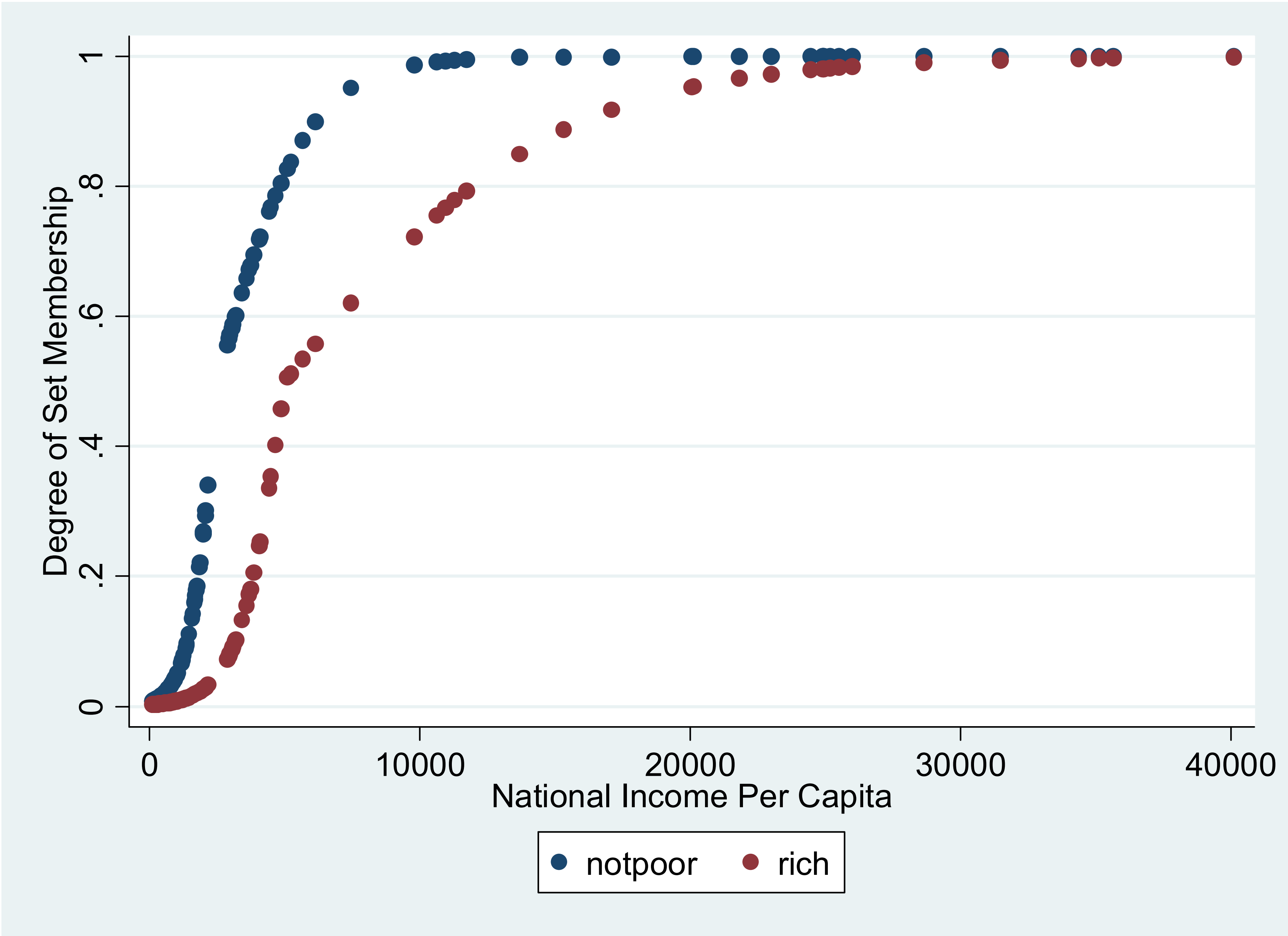
FUZZY MEMBERSHIP IN THE SET OF "POOR" COUNTRIES"

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

Plot of Degree of Membership in the Set of Poor Countries Against National Income Per Capita



Comparison of the Calibration of "Rich" Versus "Not-Poor"



Fuzzy membership scores:

. . .should be true to the label of the set. Different labels imply different calibration schemes. For example, it is much easier to have strong membership in the set of “at least upper-middle class” individuals than it is to have strong membership in the set of “rich” individuals.

. . . do not need to be symmetrically coded. For example, there might be a big gap between full membership (e.g., in the set college educated individuals) and the second highest fuzzy score (e.g., for individuals with three years of college).

. . . ideally, should range from 0 (or scores close to 0, e.g., < 0.20) to 1.0 (or scores close to 1.0, e.g., > 0.80). Avoid using fuzzy sets with all membership scores above 0.5 or all scores below 0.5.

. . . can be “double coded.” For example, you might want to compare the impact of membership in the set of “rich” with the impact of membership in the set of “not-poor” countries. (The latter set is much more inclusive.) These different calibrations often have direct theoretical relevance.

. . . are interpretive and often there is a “back-and-forth” between cases and data before the calibration of membership scores is complete.

Operations on Fuzzy Sets

1. Set union: $A + B = \max(A_i, B_i)$

2. Set intersection: $A*B = \min(A_i, B_i)$

3. Set negation: $\sim A_i = 1 - A_i$

4. Subset relation: $X_i \leq Y_i$ (set X is a subset of set Y)

5. Superset relation: $X_i \geq Y_i$ (set X is a superset of set Y)

6. Consistency of set X as a subset of set Y: $\sum \min(X_i, Y_i) / \sum (X_i)$

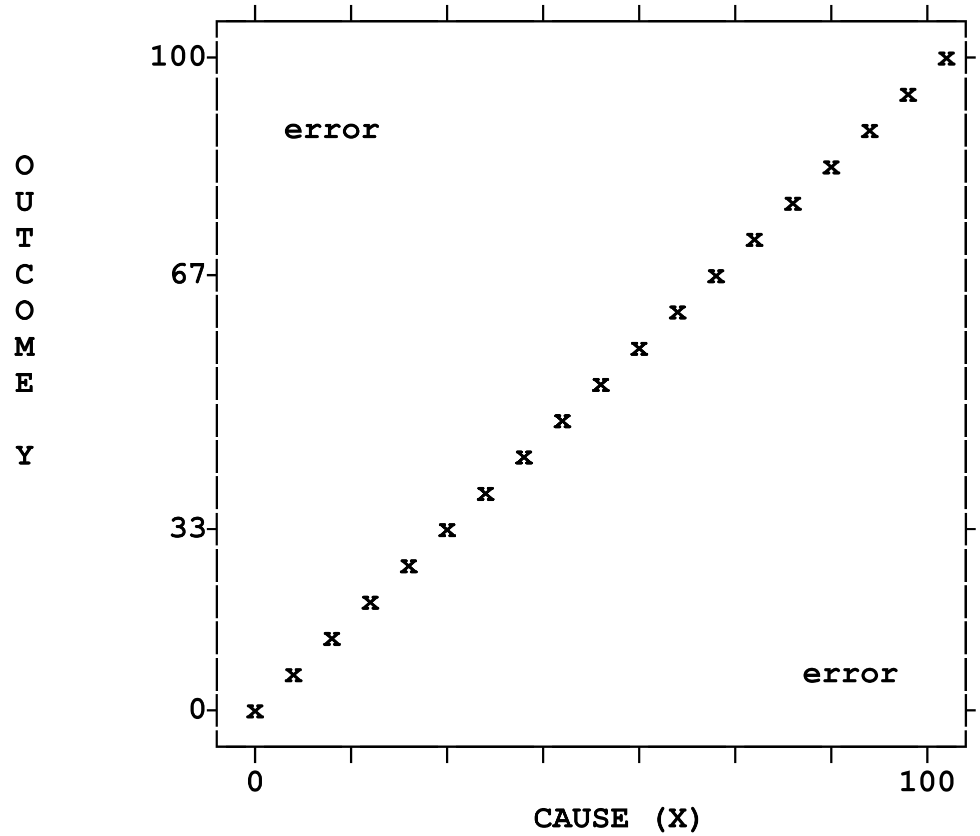
7. Coverage of set Y by subset X: $\sum \min(X_i, Y_i) / \sum (Y_i)$

8. Coincidence of sets A and B: $\sum \min(A_i, B_i) / \sum \max(A_i, B_i)$

9. Degree of membership in a combination of sets (multiple set intersection):
   membership = $\min(A_i, B_i, C_i, D_i, E_i)$

10. Degree of membership in alternate sets (multiple set union):
    membership = $\max(A_i, B_i, C_i, D_i, E_i)$

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 a conventional 2X2 crosstabulation are 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).

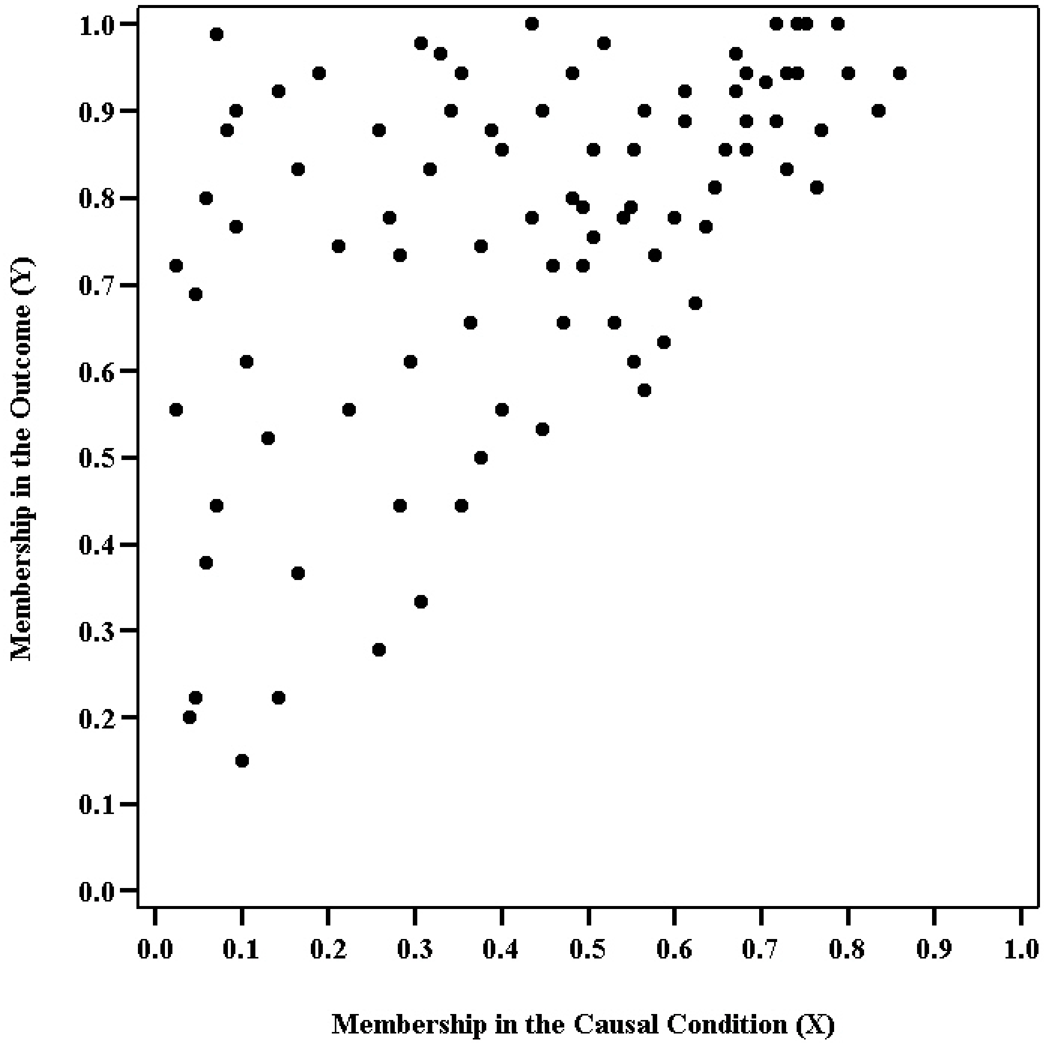


Figure 1: Fuzzy Subset Relation Consistent with Sufficiency

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 or that instances of X share outcome 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.

Note that X could be a single causal condition or a combination of conditions. If the latter, membership in the combination is the minimum of the membership scores in the fuzzy sets that are included in the combination (fuzzy set intersection).

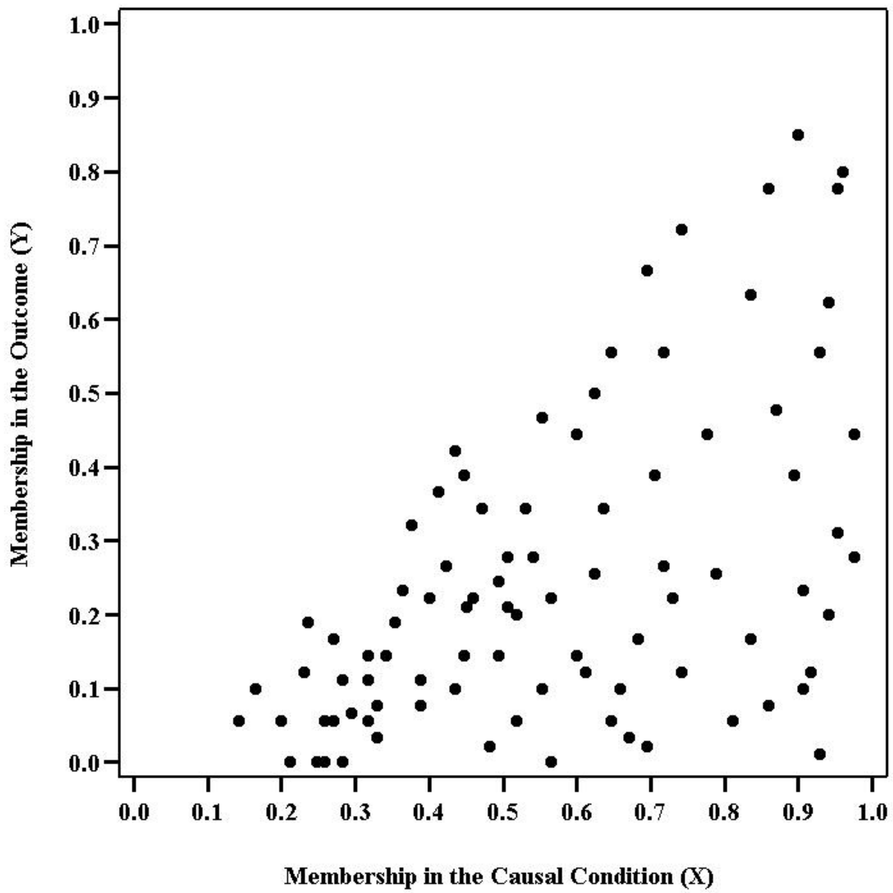


Figure 2: Fuzzy Subset Relation Consistent with Necessity

This plot illustrates the characteristic lower-triangular plot indicating the fuzzy superset relation: X ≥ 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 or simply that instances of Y share X as an antecedent condition.

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, by contrast, 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).

Again note that X could be a single causal condition or a combination of conditions. If the latter, membership in the combination is the minimum of the membership scores in the fuzzy sets that are included in the combination (fuzzy set intersection).

USING LOGICAL “AND” TO ESTABLISH A SET RELATION CONSISTENT WITH SUFFICIENCY

When elaborating the subset relation with fuzzy sets, with the cause or causal combination as a subset of the outcome, the goal is to “move” cases to the left side of the main diagonal of the scatterplot (i.e., above it).

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 their 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 and logically that membership in A*B*C [min(A,B,C)] is less than or equal to membership in A*B [min(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.

USING LOGICAL "OR" TO ESTABLISH A SET RELATION CONSISTENT WITH NECESSITY

When elaborating the superset relation with fuzzy sets, with the cause or causal combination as a superset of the outcome, the goal is to "move" cases to the right side of the main diagonal of the scatterplot (i.e., below it).

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 their 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 sets that are joined as the membership of a case in their union. It follows mathematically and logically that membership in A+B+C [max(A,B,C)] is greater than or equal to membership in A+B [max(A,B)].

Thus, the elaboration of a superset relation through substitutable conditions raises the X values and thus may move cases toward the right side of the diagonal.