CONFOUNDER-INDUCED SPURIOSITY AND REVERSAL: ALGEBRAIC CONDITIONS USING A NON-INTERACTIVE MODEL FOR BINARY DATA

Milo Schield, Augsburg College and Tom Burnham, Cognitive Consulting
Dept. of Business Administration. 2211 Riverside Drive. Minneapolis, MN 55454

Abstract: Defining conditions are obtained under which a binary confounder will nullify (render spurious) or reverse an association between binary variables when using a non-interactive (NI) linear OLS regression model. These defining conditions are used to derive necessary conditions for NI spuriosity and reversal. These necessary conditions include generalizations of those obtained by Cornfield and Gastwirth. Cornfield's "no effect" condition for spuriosity is found to be a special case of NI spuriosity. The reversal which occurs in Simpson's paradox is found to be a special case of NI reversal. Simple tests are obtained to infer whether an association will be increased, decreased or reversed after controlling for a confounder.

1. BACKGROUND

This paper deals with confounder-induced spuriosity.[1] An association between two variables is *confounded* by a third if the third has an influence on their association. An association is *spurious* – of no effect – if it vanishes after taking a confounder into account. Let *E* be a binary effect and let *A* and *B* be binary predictors. The goal of this paper is to identify the conditions when the association between *A* and *E* becomes spurious or reverses (changes sign) after taking into account a confounder, *B*, using a non-interactive model.

2. NOTATION

The variable name is used to indicate the values (e.g., *A* and *non-A*). *A'* designates non-*A*. If *E* is cancer and *A* is smoker, then $P(E|A')$ is the prevalence of cancer for non-smokers.[2] In order to study differences between, and ratios of, prevalences, this notation is used:

1. $DP(Y:X) \equiv P(Y|X)-P(Y|X')$,
2. $RP(Y:X) \equiv P(Y|X)/P(Y|X')$, $XRP(Y:X) \equiv RP(Y:X)-1$,
3. $AFP(Y:X) \equiv DP(Y:X) \cdot P(X)/P(Y)$.

The colon indicates that the following value and its complement are involved. Consider cancer (*E*), smoking (*A*) and a cancer gene (*B*). *DP(B:A)* is the differential prevalence of the cancer gene for smokers vs. non-smokers. *RP(E:A)* is the relative prevalence, *XRP(E:A)* is the excess relative prevalence, of cancer for smokers vs. non-smokers. *AFP(E:A)* is the fraction of cancer cases in the population that are attributed to smoking.

The selection of *A* vs. *A'*, and of *B* vs. *B'* is arbitrary. This paper assumes they are selected so $DP(E:A) > 0$ and $DP(E:B) > 0$.[3] These selections do not determine whether *DP(B:A)* is positive in general.

3. SPURIOSITY AS "NO EFFECT"

The first categorical criterion for spuriosity arose in the argument about whether smoking causes lung cancer. A clear association had been demonstrated. But was smoking the *direct cause* of cancer or was the association spurious – due to some confounder? In 1958, Fisher, a leading statistician and a smoker, argued that the smoking-cancer association might be confounded by genetics. He found an association for twins between the degree of twinship (identical or fraternal) and smoking preference. To reply, Cornfield modeled spuriosity by assuming smoking (*A*) had "no effect":

4. $P(E|B,A) = P(E|B,A') = P(E|B)$,
5. $P(E|B',A) = P(E|B',A') = P(E|B')$.

We call these conditions "**cross-A rate equalities**" because the rates are equal across A (conditionally independent of A). In equations derived from 4 and 5, *B* is replaced by *b* to indicate these equalities. These restrictions are not on B, but on $P(E|B)$ and $P(E|B')$. Cornfield derived a variation of this equation:[4]

6. $$RP(E:A) = \frac{[P(b\mid A) \bullet XRP(E:b)]+1}{[P(b\mid A') \bullet XRP(E:b)]+1}.$$

From his variation, Cornfield derived this condition:[5]

7. $RP(E:A) < RP(b:A)$.

Cornfield et al. (1959) replied to Fisher (italics added):

> "Thus, if cigarette smokers have 9 times the risk of nonsmokers for developing lung cancer [$RP(E:A)$=9], and this is *not because cigarette smoke is a causal agent*, but only because cigarette smokers produce hormone X, then the proportion of hormone-X producers among cigarette smokers must be at least 9 times greater than that of non-smokers [$RP(b:A)$>9]."[6]

Fisher never replied. Statisticians then asserted that smoking caused cancer using observational data.

Using the cross-A rate equality conditions (Eq. 4 and 5), Cornfield also derived a difference equality:

8. $DP(E:A) = DP(E:b) \cdot DP(b:A)$.

---

[1] A spurious association can also be chance-based: due to sampling variability when there is no association in the population.

[2] Note that $P(X)$ signifies prevalence or percentage – not probability.

[3] If $DP(E:A) = 0$ then reversal is not meaningful. If $DP(E:B) = 0$ or $DP(B:A) = 0$, then spuriosity and reversal are impossible (Eq. 25).

[4] In Eq. 6, $P(b|A) > P(b|A')$ since $RP(E:A) > 1$ and $XRP(E:b) > 0$.

[5] Eq. 6 has the form $Z = (U \cdot X + 1)/(V \cdot X + 1)$ with $U>0$, $V>0$ and $X>0$. So $Z=(U/V)(X+1/U)/(X+1/V)$. Footnote 4: $P(b|A) > P(b|A')$. So $U > V$, $1/U < 1/V$, $(X+1/U) < (X+1/V)$ and $(X+1/U)/(X+1/V) < 1$. $U/V = RP(b:A)$, so $Z < U/V$ and $RP(E:A) < RP(b:A)$.

[6] Appendix A of Schield (1999) replicates Cornfield's derivation.

Thus, if the association between smoking and cancer is spurious, then the differential cancer prevalence for smokers vs. non-smokers, *DP(E:A),* must equal the differential cancer prevalence for cancer-gene carriers vs. non-carriers, *DP(E:b),* times the differential cancer-gene prevalence for smokers vs. non-smokers, *DP(b:A)*. Cornfield did not see this as useful.[7]

Gastwirth (1988) used Cornfield's "no effect" assumption to derive another expression for spuriosity:

9. $$RP(b:A) = RP(E:A) + \frac{RP(E:A)-1}{P(b \mid A')[RP(E:b)-1]}.$$

Cornfield's condition follows from this since the fraction is positive. From a form of Eq. 6, Gastwirth derived a second necessary condition:[8,9]

10. $RP(E:A) \leq RP(E:b)$.

If the smoking-cancer association is due to a gene, this condition means that the relative prevalence of cancer among smokers vs. non-smokers [*RP(E:A)*] must be less than or equal to the relative prevalence of cancer among those with vs. without the gene [*RP(E:b)*].

4. NON-INTERACTIVE SPURIOSITY

In the following models, the values of the variables are treated as continuous. Rather than use new notation, we ask readers to recognize that *E*, *A* and *B* can be continuous in Eq. 11-12, 14-16 and figures 1-4, 7 and 8.

Consider modeling *E* on two continuous predictors *A* and *B*. When the regression coefficient between *A* and *E* is zero, that relationship is said to be 'spurious' with respect to *B*. When the model is linear and non-interactive (NI), the regression coefficient relating *E* and *A* is proportional to $r_{AE,B}$, the partial correlation coefficient between *A* and *E* after controlling for *B*:[10,11]

11. $$r_{AE,B} = (r_{AE} - r_{AB} \cdot r_{BE}) / \sqrt{(1-r_{AB}^2)(1-r_{BE}^2)}.$$

NI spuriosity occurs when $r_{AE,B} = 0$. This implies that:

12. $r_{AE} = r_{AB} \cdot r_{BE}$.

Schield (1999) applied this well-known condition for spuriosity to binary data and obtained this condition:

13. $DP(E:A) = DP(E:B) \cdot DP(B:A)$.

This condition (Eq. 13) is similar to the condition in Eq. 8, but without the cross-A rate equality assumption. *DP(B:A)* > 0 for NI spuriosity (since *DP(E:A)* > 0 and *DP(E:B)* > 0) and for NI reversal (defined in Section 6) as proven in section 8 after Eq. 24.

[7] "if the absolute difference, R1 - R2, is used, the relationship, R1-R2 = (r1-r2)(p1-p2), leads to no useful conclusion about p1-p2."

[8] Eq. 6 has the form, $Z = [U(Y-1)+1]/[V(Y-1)+1]$. U>0, V>0, Y>1. Since [V(Y-1)+1] > 1, [U(Y-1)+1]/[V(Y-1)+1] < [U(Y-1)+1]. So, $Z < [U \cdot (Y-1)+1]$. Since $U < 1$, $Z < (Y-1)+1$. So $Z < Y = RP(E:b)$.

[9] Gastwirth (1988) attributed this condition to Cornfield. Since Gastwirth first derived it, we call it the Gastwirth-Cornfield condition.

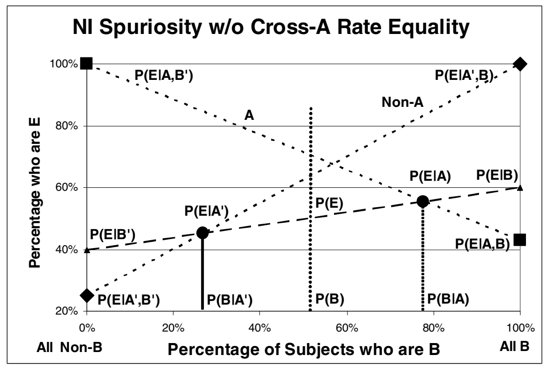
[10] Note: r_{AE} is the Pearson correlation coefficient between *E* and *A*.

[11] If $r_{BE} = 0$ then $|r_{AE,B}| > |r_{AE}|$. So, the association between *A* and *E* can not be nullified or reversed by such a confounder.

5. CROSS-A VS. NI SPURIOSITY

Since both the cross-A rate equality condition and the NI model give similar results (Eq. 8 and Eq. 13), it may be worth explicating their difference. The difference equation (Eq. 13) can be written as equal slopes: $\Delta Y/\Delta X$ = *DP*(*E:A*)/*DP*(*B:A*) = *DP*(*E:B*)/(1-0). For other forms see Equations A9 in Appendix A. Figure 1 shows data that satisfies this slope condition.

Figure 1: Non-Interactive (NI) Spuriosity[12]



In cross-A rate equality, *P*(*E*|*A',B*)=*P*(*E*|*A,B*)=*P*(*E*|*B*) and *P*(*E*|*A',B'*)=*P*(*E*|*A,B'*)=*P*(*E*|*B'*). So Figure 1 does not involve cross-A rate equality. *P*(*E*|*B*) is always a weighted average of two rates: *P*(*E*|*A,B*) and *P*(*E*|*A',B*). For cross-A rate equality, these rates are equal, so the weights don't matter. For non-interactive spuriosity, these rates can be unequal so the weights do matter.

6. NON-INTERACTIVE REVERSAL

Non-interactive (NI) reversal is readily seen using the regression approach presented by Wonnacott and Wonnacott (1990, Appendix 13-5). A regression model generates a line, *E*(*A*) or *B*(*A*), or a surface, *E*(*A,B*):

14. $E(A) = b_0(E|A) + b_1(E|A) \cdot A$,

15. $E(A,B) = b_0(E|A,B) + b_1(E|A,B) \cdot A + b_2(E|A,B) \cdot B$,

16. $B(A) = b_0(B|A) + b_1(B|A) \cdot A$.

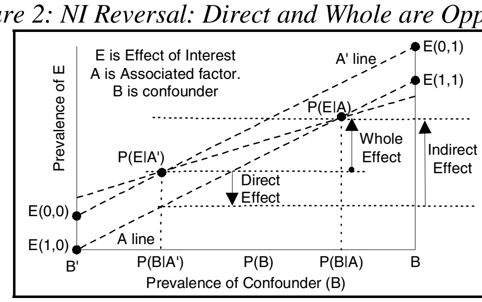
They showed these four slopes are related as follows:

17. $b_1(E|A) = b_1(E|A,B) + [b_2(E|A,B) \cdot b_1(B|A)]$.

18. *whole effect = direct effect + indirect effect.*

An NI model with two binary predictors generates a surface, *E*(*A,B*), that forms two parallel lines: the A′ line, *E*(*A=0,B*), and the A line, *E*(*A=1,B*). See Figure 2.

Figure 2: NI Reversal: Direct and Whole are Opposite



[12] $P(E|A',B') = 2/8$, $P(E|A',B) = 3/3$, $P(E|A,B') = 2/2$, $P(E|A,B) = 3/7$. $P(A',B') = 8/20$, $P(A',B) = 3/20$, $P(A,B') = 2/20$ and $P(A,B) = 7/20$.

The A' line always runs through $P(E|A')$; the A line always runs through $P(E|A)$.

The **whole effect** of A on E, $b_1(E|A)$, is $DP(E:A)$ while $b_1(B|A)$ is $DP(B:A)$. The **direct effect** of A on E is the vertical distance between the two lines.

Non-interactive (NI) reversal of the association between A and E occurs when the signs of their coefficients are opposite in the one and two factor models:

19. $b_1(E|A)\cdot b_1(E|A,B) < 0$.

Thus, NI reversal occurs when the sign of the whole effect is opposite the sign of the direct effect. Since $DP(E:A) > 0$ the whole effect is positive and the direct effect is negative. If $DP(B:A) > 0$, the A line lies beneath the A' line: a geometric condition for NI reversal.

7. DEFINING AND NECESSARY CONDITIONS

Non-interactive (NI) spuriosity is also defined by:

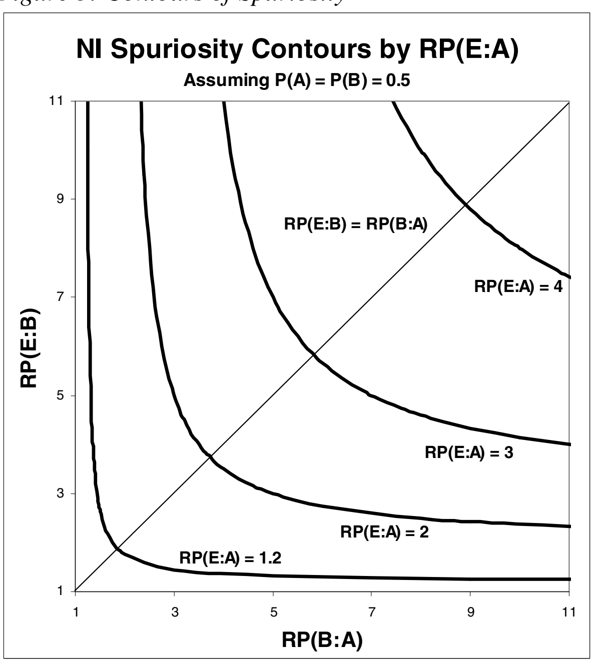
20. $b_1(E|A,B) = 0$.

Although correlation (Eq. 11 and 12) is a primary defining condition, Eq. 20 follows from their direct relationship. Appendix A contains consequences of Eq. 20. Appendices B through E give details on NI modeling.

If the association between A and E is NI spurious, then $b_2(E|A,B) = DP(E:B)$ as shown in footnote 39, the direct effect is zero, the whole effect equals the indirect effect, and we obtain Eq. 13.

$RP(E:B)$ and $RP(B:A)$ are inversely related under NI spuriosity (as are r_{BE} and r_{BA} in Eq. 12). Figure 3 displays this relationship using Eq. A4b in Appendix A.

Figure 3: Contours of Spuriosity



NI spuriosity and NI reversal are closely related. The defining condition for NI spuriosity (Eq. 20) is a boundary of the defining condition for NI reversal (Eq. 19). Since $b_1(E|A) = DP(E:A)$ and since we are assuming that $DP(E:A) > 0$, we can state the defining condition for NI reversal as:

21. $b_1(E|A,B) < 0$.

When little is known about the confounder, necessary conditions can be weaker but more useful. Any condition that is necessary for NI spuriosity, $b_1(E|A,B) = 0$, and NI reversal, $b_1(E|A,B) < 0$, is necessary for either.[13] One condition necessary for both is the combination of the defining conditions for each:

22. $b_1(E|A,B) \le 0$.

Any condition necessary for this is necessary for each.[14]

8. JOINT NECESSARY CONDITIONS

From Eq. E1 in Appendix E, it follows that:[15]

23. $b_1(E|A,B) = K1\,[DP(E:A) - DP(B:A)\cdot DP(E:B)]$.

Since $K1 > 0$, combining the joint condition (Eq. 22) with this form of b_1 gives this necessary condition:

24. $DP(E:A) \le DP(B:A)\cdot DP(E:B)$.

Since $DP(E:A) > 0$ and $DP(E:B) > 0$, it follows that $DP(B:A) > 0$, so $RP(B:A) > 1$, for both NI spuriosity and NI reversal. Since $0 < DP \le 1$,[16]

25. $DP(E:A) \le DP(B:A)$ and $DP(E:A) \le DP(E:B)$.

Similarly structured relations involving correlation coefficients are obtained from Eq. 11.[17]

From Eq. E2 in Appendix E, it follows that:

26. $b_1(E|A,B) = K2[AFP(E:A) - AFP(B:A)\cdot AFP(E:B)]$.

Since $K2 > 0$, combining the joint condition (Eq. 22) with this form of b_1 gives this necessary condition:

27. $AFP(E:A) \le AFP(B:A)\cdot AFP(E:B)$.

AFP is the fraction of E attributable to A in the population. Since $0 < AFP < 1$,[18]

28. $AFP(E:A) < AFP(E:B)$; $AFP(E:A) < AFP(B:A)$.

From Eq. E3 in Appendix E, it follows that:

29. $$b_1(E \mid A,B) = K3\{XRP(E:A)[P(B \mid A')\bullet XRP(E:B)+1] - [P(B \mid A')\bullet XRP(B:A)\bullet XRP(E:B)]\}.$$

Since $K3 > 0$, combining the joint condition (Eq. 22) with this form of b_1 gives this necessary condition:

30. $$XRP(E:A) \le \frac{XRP(B:A)\bullet P(B \mid A')\bullet XRP(E:B)}{1+[P(B \mid A')\bullet XRP(E:B)]}.$$

---

[13] Necessary conditions exist for one that are not necessary for the other. $RP(B:A) < P(E|A) / [P(E|A')-P(E|B')]$ (from Eq. A12) is necessary for NI spuriosity, but not for all NI reversals.

[14] If a joint necessary condition is $L \le R$ then an increase in R or decrease in L makes NewL < NewR a necessary condition for both. If a necessary condition is false, then the conclusion is false.

[15] Recall that the whole effect is $b_1(E|A) = DP(E:A)$. If $K1 = 1$, the indirect effect is $DP(B:A)\cdot DP(E:B)$, but this is a degenerate case.

[16] If $DP(B:A) = 1$, we have collinearity: a non-useful degenerate case.

[17] $DP(E:A) > 0$ and $DP(E:B) > 0$, so $r_{AE} > 0$ and $r_{BE} > 0$. Since $b_1(E|A,B)$ is proportional to $r_{AE,B}$, applying Eq. 22 to Eq. 11 gives $r_{AE} \le r_{AB}\cdot r_{BE}$ as a necessary condition for NI spuriosity and reversal. So $r_{AE} \le r_{AB}\cdot r_{BE}$, $r_{AE} \le r_{AB}$, and $r_{AE} \le r_{BE}$ are necessary for NI spuriosity and reversal. These are analogs of Eq. 24 and 25.

[18] $DP(E:A)=AFP(E:A)\cdot P(E)/P(A)$. $DP(E:A)>0$ implies $AFP(E:A)>0$.

The denominator is more than 1; the product of the first two factors in the numerator is less than 1.[19] Replacing both with 1 gives a necessary condition that is a generalization of the Gastwirth-Cornfield condition (Eq. 10):

31. $XRP(E{:}A) < XRP(E{:}B)$, and $RP(E{:}A) < RP(E{:}B)$.

In Eq. 30, the denominator is greater than 1, so the inequality remains if we replace it with 1. This generates:

32. $XRP(E{:}A) < XRP(B{:}A)\cdot P(B|A')\cdot XRP(E{:}B)$.

If $XRP(B{:}A)\cdot P(B|A') < 1$, Eq. 32 is stronger than Eq. 31. If $XRP(E{:}B)\cdot P(B|A') < 1$, Eq. 32 is stronger than Eq. 35.

From Eq. E4 in Appendix E, it follows that:

33. $b_1(E|A,B) = K4\{[P(B)\cdot XRP(B{:}A)\cdot XRP(E{:}B)] + XRP(E{:}A)[P(A)\cdot XRP(B{:}A)+P(B)\cdot XRP(E{:}B)+1]\}$.

Since $K4 > 0$, combining the joint condition (Eq. 22) with this form of b_1 gives this necessary condition:

34. $$XRP(E:A) \leq \frac{P(B)\bullet XRP(B:A)\bullet XRP(E:B)}{P(B)\bullet XRP(E:B) + P(A)\bullet XRP(B:A)+1}.$$

Since the items being added in the denominator are positive, we can retain the inequality by retaining any one of them. Doing this from left to right gives these three necessary conditions:

35. $XRP(E{:}A) < XRP(B{:}A)$,
36. $XRP(E{:}A) < [P(B)/P(A)]\cdot XRP(E{:}B)$,
37. $XRP(E{:}A) < P(B)\cdot XRP(B{:}A)\cdot XRP(E{:}B)$.

Eq. 35 is a generalization of Cornfield's condition (Eq. 7). Eq. 36 is more restrictive than the generalized Gastwirth-Cornfield condition (Eq. 31) if $P(B) < P(A)$. Eq. 37 is less restrictive than Eq. 32 but might be more useful as is the following: [20,21]

38. $XRP(E{:}A) < XRP(B{:}A)\cdot XRP(E{:}B)$.

For the case of smoking and cancer, the generalization (Eq. 31) of the Gastwirth-Cornfield condition means that if this association were spurious and $RP(E{:}A)$ were 9, then $RP(E{:}B)$ must be greater than 9 for a hypothetical genetic confounder. But if the prevalence of such a genetic confounder, $P(B)$, was 10%, and the smoker prevalence, $P(A)$, was 40%, then this new condition (Eq. 36) would require $RP(E{:}B) > 33$.

9. "NO EFFECT" SPURIOSITY

Under NI spuriosity, the two cross-A rate differences, $DP(E{:}A|B)$ and $DP(E{:}A|B')$[22], must either be opposite in sign (Figure 1) or zero (cross-A rate equality).

In Appendix D, it is shown that any instance of cross-A rate equality must involve NI spuriosity. Since Figure 1 is an example of NI spuriosity which does not involve cross-A rate equality, we infer that cross-A rate equality is a special case of NI spuriosity.

[19] $[XRP(B{:}A)\cdot P(B|A')] = [P(B|A) - P(B|A')] < P(B|A) < 1$.
[20] This is more restrictive than Eq. 31 & 35 if both $XPRs < 1$. It is more useful than Eq. 32 or 36 if $P(B|A)$ and $P(B|A')$ are unknown.
[21] $RP(E{:}A)-1 < [RP(B{:}A)-1][RP(E{:}B)-1] < [RP(B{:}A)\cdot RP(E{:}B) - RP(B{:}A) - RP(E{:}B) + 1] < RP(B{:}A)\cdot RP(E{:}B) - 1$.
[22] $DP(Z{:}X|Y') \equiv [P(Z|X,Y') - P(Z|X',Y')]$ is analogous to Eq. 1.

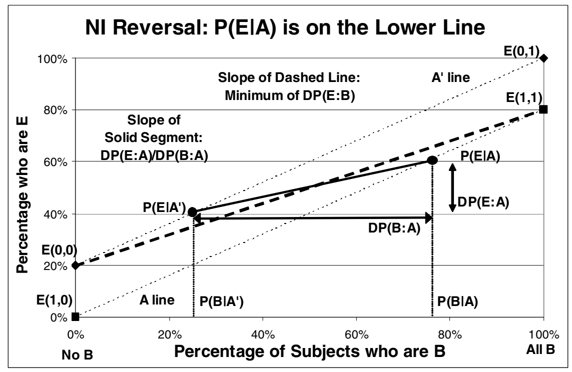
10. GEOMETRY OF NI REVERSAL

Eq. 21 gives a defining condition for NI reversal. Using Eq. 23 with Eq. 21 gives this form:

39. $DP(E{:}A) / DP(B{:}A) < DP(E{:}B)$.

Figure 4 illustrates this condition graphically. The light dotted lines are the edges of the $E(A,B)$ surface for A and A' where the A line lies below the A' line. $P(E|B)$ is between $E(0,1)$ and $E(1,1)$; $P(E|B')$ is between $E(0,0)$ and $E(1,0)$. See Eq. D6. $DP(E{:}A)/DP(B{:}A)$ is the slope of the dark solid segment. The slope of the dashed line, $[E(1,1)-E(0,0)]/1$, is the maximum of $DP(E{:}A)/DP(B{:}A)$ and the minimum of $DP(E{:}B)/1$.[23]

Figure 4: Geometric Condition for NI Reversal



A geometric condition for NI reversal is that the A line lies below the A' line so $P(E|A)$ lies on the lower line.

11. SIMPSON'S REVERSAL

Simpson's Paradox exists when the sign of association in *each* sub-group (B and B') is opposite the sign in the composite group. We define **Simpson's reversal** as the reversal occurring in Simpson's Paradox: [24]

40. $DP(E{:}A|B) < 0$, $DP(E{:}A|B') < 0$ when $DP(E{:}A)>0$.

Not all NI reversals involve a Simpson's reversal. Figure 1 illustrates an NI reversal but not all the signs of the sub-group differences are opposite that in the composite: $DP(E{:}A|B) < 0$ but $DP(E{:}A|B') > 0$.

Simpson's reversal cannot occur without NI reversal as shown using this identity (Eq. B8 in Appendix B):

41. $DP(E{:}A) = DP(B{:}A)\cdot DP(E{:}B) + X$,
42. $X = [P(B|A)\cdot DP(E{:}A|B)\cdot P(B|A')/P(B)] + [P(B'|A)\cdot DP(E{:}A|B')\cdot P(B'|A')/P(B')]$.

In Eq. 41, $X < 0$ is another form of the defining condition for NI reversal (see Eq. 39). As defined in Eq. 40, a Simpson's reversal is sufficient to make $X < 0$ in Eq. 41. So, all instances of Simpson's reversal must involve an NI reversal. But not vice versa since a Simpson's reversal is not necessary for $X < 0$ in Eq. 42.

[23] The maximum of $DP(E{:}A)/DP(B{:}A)$ and minimum of $DP(E{:}B)/1$ are achieved simultaneously only under NI spuriosity.
[24] If the underlying rates were coplanar with cross-A rate difference equality, $DP(E{:}A|B) = DP(E{:}A|B')$, then $E(1,1) = P(E|A,B)$, etc., See Eq. D2e. If so, Figure 4 would illustrate a Simpson's reversal. E.g., $DP(E{:}A|B) = [P(E|A,B) - P(E|A',B)] = [E(1,1) - E(0,1)] < 0$.

12. INFERENCES

The influence of a confounder, B, on an observed association between A and E can be inferred without doing the regression provided one has information on comparisons of single-predictor prevalences: $P(X|Y)$. Assume as usual that values of A and B are selected so $DP(E{:}A) > 0$ and $DP(E{:}B) > 0$. We describe three cases: (1) given the signs of three comparisons, (2) given three relative differences, and (3) given three simple differences.

#1: Direction of Change

Since $b_1(E|A) = DP(E{:}A)$, Eq. 23 can be rewritten as:

43. $b_1(E|A,B) = K1[b_1(E|A) - DP(B{:}A) \cdot DP(E{:}B)]$.

The direction of change in the association between A and E can be inferred from the sign of $DP(B{:}A)$:

44. Decrease: $b_1(E|A,B) < b_1(E|A)$ if $DP(B{:}A) > 0$.

45. Increase: $b_1(E|A,B) > b_1(E|A)$ if $DP(B{:}A) < 0$.

Since XRP has the same sign as DP, the sign of $XRP(B{:}A)$ can be used to infer the direction of change.

#2: Non-Reversal[25,26]

If $XRP(B{:}A) > 0$, then $b_1(E|A,B) < b_1(E|A)$. In this case, an NI reversal, $b_1(E|A,B) < 0$, is precluded if any of the following are true:

46. $XRP(E{:}A) > XRP(E{:}B)$, $XRP(E{:}A) > XRP(B{:}A)$, or $XRP(E{:}A) > XRP(B{:}A) \cdot XRP(E{:}B)$.

Eq. 46 follows from Eq. 31, 35 and 38 respectively. If all of the known elements of Eq. 46 are false, then an NI reversal is not precluded.

#3: Reversal

When rearranged, Eq. 39 gives this form of the defining condition for NI reversal:

47. $DP(E{:}A) < DP(B{:}A) \cdot DP(E{:}B)$.

If Eq. 47 is true, then an NI reversal holds after taking the confounder into account; otherwise it does not.

13. AN EXAMPLE

The relevant outcome (E) is death, A is hospital (city vs. rural), and B is patient condition (poor vs. good).

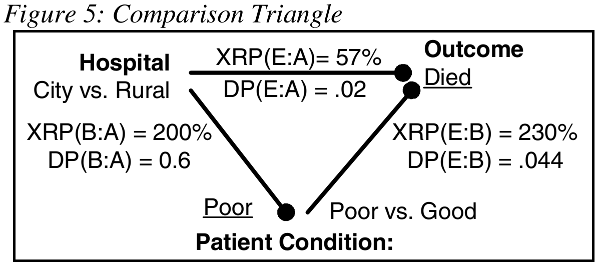
(#1) Suppose qualitative comparisons are obtained as follows. Death is more prevalent among patients at city hospitals that among those at rural hospitals; death is more prevalent among patients admitted in poor condition than among those admitted in good condition; and admission in poor condition is more prevalent among patients at city hospitals than among those at rural hospitals. It follows that the association between city hospitals and higher death rates is decreased after controlling for patient condition because all three DPs or XRPs are positive.

(#2) Suppose percentage comparisons are obtained as follows. Death is 57% more prevalent among patients at city hospitals than among those at rural hospitals, so $XRP(E{:}A) = 0.57$. Death is 230% more prevalent for patients admitted in poor condition than for patients admitted in good condition, so $XRP(E{:}B) = 2.3$. And admission in poor condition is 200% more prevalent among patients at city hospitals than among patients at rural hospitals, so $XRP(B{:}A) = 2.0$. As in #1, the association between city hospitals and higher death rate is decreased by taking into account patient condition. In addition, it follows that a reversal of the association is not precluded, because $XRP(E{:}B)$, $XRP(B{:}A)$, and $XRP(E{:}B) \cdot XRP(B{:}A)$ are each larger than the observed difference, $XRP(E{:}A)$.[27]

(#3) Suppose percentage-point differences are obtained as follows. Death is 2 percentage points more prevalent among patients at city hospitals than among those at rural hospitals, so $DP(E{:}A) = 0.02$. Death is 4.4 percentage points more prevalent for patients admitted in poor condition than for patients admitted in good condition, so $DP(E{:}B) = 0.044$. And admission in poor condition is 60 percentage points more prevalent among patients at city hospitals than among patients at rural hospitals, so $DP(B{:}A) = 0.6$. It follows that this association between city hospitals and higher death rates is reversed by taking patient condition into account, because the product of the two confounder-related simple differences, 0.6 times 0.044, is greater than the observed simple difference, $DP(E{:}A) = .02$.[28]

Figure 5 summarizes these comparisons. An underscore or dot indicates the common numerator in each.

Figure 5: Comparison Triangle



Now consider similar data in which admission in poor condition is just 30 percentage points more prevalent among patients at city hospitals than among those at rural hospitals. It follows that the association between city hospitals and higher death rates is not reversed by taking into account patient condition, because the confounder linkages are not strong enough to reverse the association: $0.30 \cdot 0.044$ is less than 0.02.[28]

---

[25] Skip this step if $DP(E{:}A)$, $DP(E{:}B)$ and $DP(B{:}A)$ are available.

[26] If $DP(B{:}A)$ or $XRP(B{:}A)$ are not available, they can be derived from a number of other statistics. For example

- $P(B|A) = [P(E|A) - P(E|A,B')]/[P(E|A,B)-P(E|A,B')]$
- $P(B|A') = [P(E|A') - P(E|A',B')]/[P(E|A',B)-P(E|A',B')]$.

They can also be derived using $Phi(B,A)$, $P(B)$ and $P(A)$:

- $[DP(B{:}A)]^2 = Phi^2(B,A)\{P(B)[1-P(B)]\}/\{P(A)[1-P(A)]\}$.

[27] Note: to multiply percentages they must first be converted to fractions.

[28] When $DP(E{:}A)/DP(B{:}A) < DP(E{:}B)$, the A line is below the A' line so NI reversal will happen. When $DP(E{:}A)/DP(B{:}A) > DP(E{:}B)$, the A line is above the A' line so NI reversal is impossible.

14. CONCLUSIONS

Defining conditions for an association to be nullified (made spurious) or reversed by a confounder are derived for binary variables using a non-interactive (NI) linear regression model.

Necessary conditions for both NI spuriosity and NI reversal are derived. These include generalizations of the Cornfield and Gastwirth-Cornfield conditions. Cross-A rate equality (Cornfield's "no effect") is found to be a special case of NI spuriosity. Simpson's reversal is found to be a special case of NI reversal.

Simple tests are obtained to infer whether controlling for a confounder will increase, decrease or reverse an association. These tests require just single-predictor comparisons. They do not require any double-predictor prevalences and they do not require doing an actual regression.

Since confounding involving binary variables is a major problem in many fields and since there is no statistical test for confounder-induced spuriosity or reversal, these results may be of general use. For example, they may be useful in specifying the minimum strength needed by a confounder to nullify or reverse an association.

REFERENCES

Cornfield, J., Haenszel, W., Hammond, E., Lilienfeld, A., Shimkin, M., and Wynder, E. (1959). *Smoking and lung cancer: Recent evidence and a discussion of some questions.* J. of National Cancer Institute, 22, pp. 173-203.

Fisher, Ronald (July, 1958). Letter to the Editor, *Lung Cancer and Cigarettes.* Nature, 182. p. 108.[29]

Gastwirth, Joseph L. (1988). *Statistical Reasoning in Law and Public Policy.* pp. 296-297. Academic Press

Lazarsfeld, Paul F. "*Algebra of Dichotomous Systems*" in *Studies in Item Analysis and Prediction* edited by Herbert Solomon (1961), p. 121, Stanford Univ. Press.

Schield, Milo (1999). *Simpson's Paradox and Cornfield's Conditions.* 1999 ASA Proceedings of the Section on Statistical Education, pp. 106-111.

Schield, Milo and Thomas V.V. Burnham (2002). *Relations between Relative Risk, Phi and Measures of Necessity and Sufficiency in 2x2 Tables.* 2002 ASA Proceedings of the Section on Statistical Education.

Wonnacott, Thomas A. and Ronald Wonnacott (1990). *Introductory Statistics, 5th ed.* John Wiley & Sons.

Acknowledgments: This work was performed under a grant from the W. M. Keck Foundation to Augsburg College "to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts." Schield is at schield@augsburg.edu. This paper is posted at www.augsburg.edu/ppages/~schield.

[29] www.economics.soton.ac.uk/staff/aldrich/fisherguide/rafreader.htm

Appendix A: NON-INTERACTIVE SPURIOSITY[30]

I. THREE DOUBLE RATIOS PER EQUATION[31]

A1. $DP(E:A) = DP(B:A) \bullet DP(E:B)$

A2. $AFP(E:A) = AFP(B:A) \bullet AFP(E:B)$

A3a. $XRP(E:A) = \dfrac{XRP(B:A) \bullet P(B|A') \bullet XRP(E:B)}{1+[P(B|A') \bullet XRP(E:B)]}$

A3b. $XRP(E:B) = \dfrac{XRP(E:A)}{P(B|A')[XRP(B:A) - XRP(E:A)]}$

A3c. $XRP(B:A) = XRP(E:A)\{1+1/[P(B|A') \bullet XRP(E:B)]\}$

A3d.

$$XRP(E:B) - XRP(E:A) = \frac{\{RP(E:B) \bullet P(B|A') + [1-P(B|A)]\}XRP(E:B)}{[P(B|A') \bullet XRP(E:B)]+1}$$

A4a. $XRP(E:A) = \dfrac{P(B) \bullet XRP(B:A) \bullet XRP(E:B)}{P(A) \bullet XRP(B:A) + P(B) \bullet XRP(E:B) + 1}$

A4b. $P(B) \bullet XRP(E:B) = \dfrac{XRP(E:A)\{1+[P(A) \bullet XRP(B:A)]\}}{[XRP(B:A) - XRP(E:A)]}$

A4c. $XRP(B:A) = \dfrac{XRP(E:A)\{1+[P(B) \bullet XRP(E:B)]\}}{[P(B) \bullet XRP(E:B)] - [P(A) \bullet XRP(E:A)]}$

A4d.

$$\frac{P(A) \bullet XRP(E:A)}{P(A) \bullet XRP(E:A)+1} = \frac{P(B) \bullet XRP(E:B)}{P(B) \bullet XRP(E:B)+1} \bullet \frac{P(A) \bullet XRP(B:A)}{P(A) \bullet XRP(B:A)+1}$$

II. TWO DOUBLE RATIOS PER EQUATION

A5. $XRP(B:A) = XRP(E:A) \bullet P(E|A')/[P(E|A') - P(E|B')]$

A6a. $RP(E:A) = \dfrac{[P(B|A) \bullet XRP(E:B)]+1}{[P(B|A') \bullet XRP(E:B)]+1}$

A6b. $XRP(E:B) = \dfrac{XRP(E:A)}{P(B|A) - P(B|A') \bullet RP(E:A)}$

A7. $XRP(E:B) = \dfrac{DP(E:A)}{P(B|A) \bullet P(E|A') - P(B|A') \bullet P(E|A)}$

A8. $P(A) = \dfrac{P(E) - P(E|A')}{DP(B:A) \bullet DP(E:B)} = \dfrac{DP(E:B)[P(B) - P(B|A')]}{DP(E:A)}$

III. EQUAL SLOPES

A9a. $\dfrac{\Delta Y}{\Delta X} = \dfrac{DP(E:A)}{DP(B:A)} = \dfrac{DP(E:B)}{(1-0)} = \dfrac{P(E|A) - P(E)}{P(B|A) - P(B)} = \dfrac{P(E|B) - P(E)}{1 - P(B)}$

A9b. $\dfrac{\Delta Y}{\Delta X} = \dfrac{P(E|A') - P(E|B')}{P(B|A')} = \dfrac{[P(E|A) - P(E|B')]}{P(B|A)}$

A9c. $\dfrac{\Delta Y}{\Delta X} = \dfrac{P(E|B) - P(E|A')}{1 - P(B|A')} = \dfrac{P(E|B) - P(E|A)}{1 - P(B|A)}$

A9d. $\dfrac{\Delta Y}{\Delta X} = \dfrac{P(E) - P(E|A')}{P(B) - P(B|A')} = \dfrac{P(E) - P(E|B')}{P(B)}$

IV. OTHER CONDITIONS (NOT SHOWN ABOVE)

A10. $P(E|A) = P(E|B') + P(B|A) \bullet DP(E:B)$

A11. $P(E|A') = P(E|B') + P(B|A') \bullet DP(E:B)$

A12. $RP(B:A) = [P(E|A) - P(E|B')]/[P(E|A') - P(E|B')]$

A13. $P(E)/P(E|A') = \dfrac{[P(E)/P(E|B')][P(B)/P(B|A')]}{[P(E)/P(E|B')] + [P(B)/P(B|A')] - 1}$

A14. This equates the whole with the indirect effect.

$$DP(E:A) = P(E|B) \bullet DP(B:A) + P(E|A) \bullet P(B|A') - P(E|A') \bullet P(B|A)$$

A15. $\dfrac{P(E|A)}{P(E)} - 1 = [\dfrac{P(E|B)}{P(E)} - 1][\dfrac{P(B|A)}{P(B)} - 1][\dfrac{P(B)}{1 - P(B)}]$

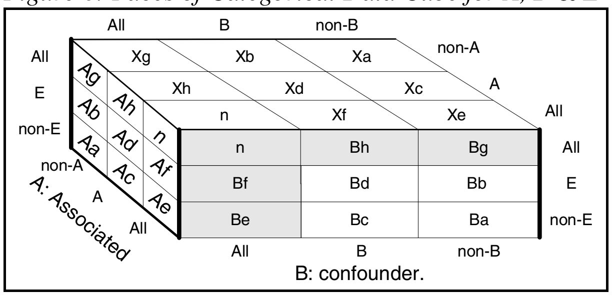
[30] $DP(E:A) = P(E) \cdot XRP(E:A)/[P(A) \cdot XRP(E:A)+1] = AFP(E:A) \cdot P(E)/P(A)$.

[31] A1, A5, A8 and A12 have two non-A ratios. All others have more.

Appendix B: DATA CUBE NOTATION

Heretofore the data values are categories (*A, A', B, B', E and E'*) and related prevalences. E.g., *P(A), P(E|A), RP(E:A)* and *DP(E:A)*. Hereafter these prevalences are given atomic symbols (proper names). Figure 6 shows the faces of the categorical cube for binary variables.

Figure 6: Faces of Categorical Data Cube for A, B & E



The three margin faces (*AE*, *BE* and *AB*) correspond to Tables A, B and X (in the next column). Body cells are labeled a though d; margin cells are e through h. Note: *Af=Bf*, *Xf=Bh* and *Xh=Ah* (*Ae=Be*, *Bg=Xe* and *Xg=Ag*); some margin cells are on more than one face.

Since our focus is on modeling outcome *E*, a 4th table, Table E, the center slice, is of interest. For each entity, we use the first letter of the variable name to indicate the table. E.g., *Xd* is cell d in Table X.

To focus on outcome *E*, we shift from counts to ratios. Table R is a ratio table: $R_i = E_i / X_i$.

The n data points are summarized by four rates (*Ra, Rb, Rc, Rd*) and their weights (*Xa , Xb, Xc, Xd*):[32]

B1. $Ra = P(E|A',B')$, $Rb = P(E|A',B)$.

B2. $Rc = P(E|A,B')$, $Rd = P(E|A,B)$.

Weights are shown relative to column or row totals:[33,34]

B3. $XP=P(B|A)$, $XQ = P(B|A')$, $XF=BH=P(B)$

B4. $XN=P(A|B)$, $XM=P(A|B')$, $XH=AH=P(A)$.

Two-letter names with all caps indicate single ratios. E.g., *BH = Bh/n*, *BF = Bf/n*. *AP* and *AQ* signify ratios in the exposure and non-exposure groups. *AF*, *AP*, *AQ*, *BP* and *BQ* are averages of pairs of rates (*Ra, Rb, Rc, Rd*) weighted by their counts (*Xa, Xb, Xc, Xd*):

B5. $AP = P(E|A)$, $AQ = P(E|A')$, $AF = P(E)$

B6. $BP = P(E|B)$, $BQ = P(E|B')$, $BF = P(E)$.

There are several general identities such as:

B7. $AP\text{-}AQ = (XP\text{-}XQ)(BP\text{-}BQ) + [AP - BP \cdot XP - BQ(1\text{-}XP)]/(1\text{-}AH)$,

B8. $AP\text{-}AQ = (XP\text{-}XQ)(BP\text{-}BQ) + [XP(Rd\text{-}Rb)XQ/BH + (1\text{-}XP)(Rc\text{-}Ra)(1\text{-}XQ)/(1\text{-}BH)]$,

B9. $AP\text{-}AQ=(XP\text{-}XQ)[(Rd\text{-}Rc)(1\text{-}AH) + (Rb\text{-}Ra)AH] + [(Rd\text{-}Rb)BH + (Rc\text{-}Ra)(1\text{-}BH)]$.

The following tables are obtained from the data cube in Figure 5. Tables A, B and X represent surfaces of the 3D data cube. Tables E, S and T are slices through the cube. Table R involves ratios between Tables E and X. Four letter names denote double ratios.[35]

Table A: Cross-prevalence between *A* and E

Table A	*Non-E*	*E*	TOTAL
Non-A	*Aa*	*Ab*	*Ag*
A	*Ac*	*Ad*	*Ah*
TOTAL	*Ae*	*Af*	*n*

Table B: Cross-prevalence between *B* and E

Table B	*Non-E*	*E*	TOTAL
Non-B	*Ba*	*Bb*	*Bg*
B	*Bc*	*Bd*	*Bh*
TOTAL	*Be*	*Bf=Af*	*n*

Table X: Cross-prevalence between *A* and B

Table X	*Non-B*	*B*	TOTAL
Non-A	*Xa*	*Xb*	*Xg=Ag*
A	*Xc*	*Xd*	*Xh=Ah*
TOTAL	*Xe=Bg*	*Xf=Bh*	*n*

Table E: Distribution of *E* by *A* and B.

Table E	*Non-B*	*B*	TOTAL
Non-A	*Ea*	*Eb*	*Eg=Ab*
A	*Ec*	*Ed*	*Eh=Ad*
TOTAL	*Ee=Bb*	*Ef=Bd*	*En=Af*

Table R: Rate **of *E* classified by *A* and B.**

Table R	*Non-B*	*B*	TOTAL
Non-A	*Ra=Ea/Xa*	*Rb=Eb/Xb*	*Rb=Ab/Ag*
A	*Rc=Ec/Xc*	*Rd=Ed/Xd*	*Rh=Ad/Ah*
TOTAL	*Re=Bb/Bg*	*Rf=Bd/Bh*	*Rn = Af/n*

Table T: Association of *A* and *E* for B=1.[36]

Table T	*Non-E*	*E*	TOTAL
Non-A	*Xb-Eb*	*Eb*	*Xb*
A	*Xd-Ed*	*Ed*	*Xd*
TOTAL	*Xf-Ef*	*Ef=Bd*	*Xf=Bh*

Table S: Association of *A* and *E* for B=0.

Table S	*Non-E*	*E*	TOTAL
Non-A	*Xa-Ea*	*Ea*	*Xa*
A	*Xc-Ec*	*Ec*	*Xc*
TOTAL	*Xe-Ee*	*Ee=Bb*	*Xe=Bg*

Eq. B8 is directly related to the **Lazarsfeld accounting formula.** See Lazarsfeld (1961). This paper does not include a comprehensive analysis of, or treatment for, $Q = 0$ or $P\text{-}Q = 1$.

[32] Absolute weights are *Xa, Xb, Xc* and *Xd,* where $Xa = n \cdot P(A', B')$, $Xb = n \cdot P(A', B)$, $Xc = n \cdot P(A, B')$, $Xd = n \cdot P(A, B)$.

[33] $XP=Xd/(Xc+Xd)$, $XQ=Xb/(Xa+Xb)$, $XN=Xd/(Xb+Xd)$, $XM=Xc/(Xa+Xc)$.

[34] $AH = (AF\text{-}AQ) / (AP\text{-}AQ) = (BH - XQ) / (XP - XQ)$, $BH = (AF\text{-}BQ) / (BP\text{-}BQ) = (AH - XM) / (XN - XM)$.

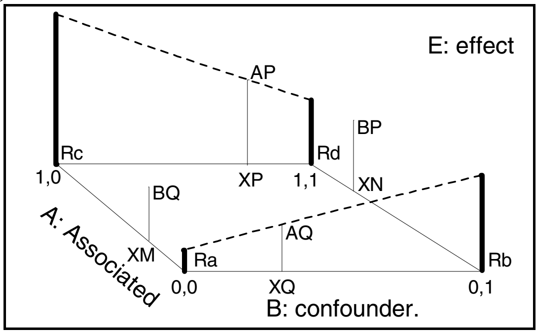
[35] $ARRE = RR(E{:}A) = AP/AQ$, $BRRE = RR(E{:}B) = BP/BQ$, $XRPB = RP(B{:}A) = XP/XQ$. See Schield and Burnham (2002).

[36] A general identity involving the S and T tables is given by: $\frac{AF(1-AH)}{AH(ARRE-1)+1} = \frac{BP(1-XN)BH}{XN(TRRE-1)+1} + \frac{BQ(1-XM)(1-BH)}{XM(SRRE-1)+1}$.

Appendix C: RATE DATA CUBE

To model this data, the values of variables A, B and E are treated as continuous. Their extreme values (A and A') are 0 and 1. See Figure 7. Location 0, 0 is A', B'. Instead of having a pair of data points (at $E = 0$ and $E = 1$) for each of the four corners, each pair is replaced by its weighted average: Ra, Rb, Rc and Rd.

Figure 7: 3D Rate Data Cube with Non-Planar Data



Noteworthy values of A are 0, XQ, $XF=AH$, XP, and 1. As shown in Figure 7, AP is a weighted average of Rc and Rd: $AP = Rc(1\text{-}XP) + Rd \cdot XP$.

Appendix D: NON-INTERACTIVE MODEL

A common linear non-interactive regression model involving two predictors is:

D1. $E(A,B) = b_0 + b_1 \cdot A + b_2 \cdot B$.

Coefficients are obtained by minimizing OLS variance. These coefficients can have many forms.

(1) One form involves rates and weights. Let b_3 indicate non-planarity where $b_3=Rd\text{-}Rc\text{-}Rb+Ra$.
Let $D = Xa[Xb(Xc+Xd)+(Xc \cdot Xd)]+(Xb \cdot Xc \cdot Xd)$,

D2a. $b_0 = Ra - (b_3 \cdot Xb \cdot Xc \cdot Xd)/D$,

D2b. $b_1 = (Rc\text{-}Ra) + [b_3 \cdot Xb(Xa+Xc)Xd]/D$,

D2c. $b_2 = (Rb\text{-}Ra)+[b_3 \cdot Xc(Xa+Xb)Xd]/D$.

Under cross-A rate equality, $Ra = Rc$ and $Rd = Rb$. So, $b_3 = 0$, $b_1 = 0$, and we have NI spuriosity.

If the data is planar, b_3 is zero, so $(Rd\text{-}Rb) = (Rc\text{-}Ra)$. So, planar data entails cross-A rate difference equality (which is different from cross-A rate equality). It also entails cross-B difference rate equality: $Rd\text{-}Rc = Rb\text{-}Ra$,

D2d. $b_0=Ra$, $b_1=Rc\text{-}Ra$, $b_2=Rb\text{-}Ra$ for planar data.

For planar data, the corners of the surface are the rates:

D2e. $E(0,0)=Ra$, $E(0,1)=Rb$, $E(1,0)=Rc$, $E(1,1)=Rd$.

(2) Another form of the coefficients involves the ratios[37] derived from the values on the faces in Figure 6. Let $D = [1 - (XN\text{-}XM)(XP\text{-}XQ)]$,

D3a. $b_0 = AF - [(AP\text{-}AQ)XM + (BP\text{-}BQ)XQ] / D$,

D3b. $b_1 = [(AP - AQ) - (BP - BQ)(XP - XQ)] / D$,

D3c. $b_2 = [(BP - BQ) - (AP - AQ)(XN - XM)] / D$.[38,39]

The following can be derived from these equations:

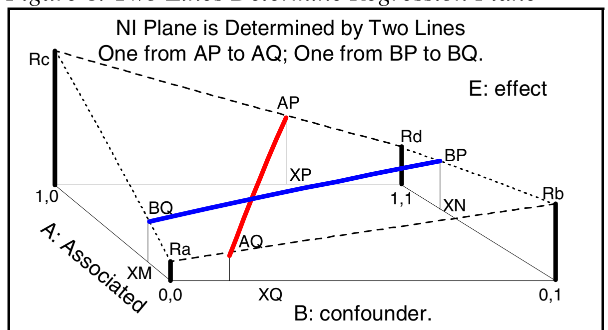
D4a. $AP = E(A=1, B=XP)$, $AQ = E(A=0, B=XQ)$,

D4b. $BP = E(A=XN, B=1)$, $BQ = E(A=XM, B=0)$,

D4c. $AF = E(AH, BH)$.

Thus the regression plane contains the lines connecting AP with AQ and BP with BQ. These lines intersect at AF. Not all ratios in categorical space lie on the surface of a given model: $Rd = P(E|A,B) \neq E(A=1,B=1)$.

Figure 8: Two Lines Determine Regression Plane



The four corners of the planar surface are:[40]

D5a. $E(0,0) = AF - [XM(AP\text{-}AQ)+XQ(BP\text{-}BQ)]/D$,

D5b. $E(0,1) = AF - [XN(AP\text{-}AQ)\text{-}(1\text{-}XQ)(BP\text{-}BQ)]/D$,

D5c. $E(1,0) = AF+[(1\text{-}XM)(AP\text{-}AQ)\text{-}XP(BP\text{-}BQ)]/D$,

D5d. $E(1,1) = AF+[(1\text{-}XN)(AP\text{-}AQ)+(1\text{-}XP)(BP\text{-}BQ)]/D$.

D6a $BP = E(0,1) + XN[E(1,1) - E(0,1)]$.

D6b. $BQ = E(0,0) + XM[E(1,0) - E(0,0)]$.

Appendix E: FORMS OF SLOPE: $b_1(E|A,B)$

The following are forms of the slope $b_1(E|A,B)$ in a non-interactive OLS regression model on binary data.

E1. $$b_1 = \frac{(AP - AQ) - (XP - XQ)(BP - BQ)}{1 - (XN - XM)(XP - XQ)}.$$ See D3b

The denominator is positive since it is $1\text{-}XPhi(B,A)^2$.

E2 $$b_1 = \frac{AF(AAFP - XAFP \bullet BAFP)}{AH[1 - (XN - XM)(XP - XQ)]}.$$

E3. A double-ratio form with XQ in numerator:

$$b_1 = \frac{AF\{(ARRE - 1)[XQ(BRRE - 1) + 1] - [XQ(XRPB - 1)(BRRE - 1)]\}}{(1 - XPhi^2)[AH(ARRE - 1) + 1][BH(BRRE - 1) + 1]}$$

E4. Double-ratio form with AH and BH in numerator:

$$b_1 = \frac{AF\{(ARRE - 1)[AH(XRPB - 1) + BH(BRRE - 1) + 1] - [BH(BRRE - 1)(XRPB - 1)]\}}{(1 - XPhi^2)[AH(XRPB - 1) + 1][AH(ARRE - 1) + 1][BH(BRRE - 1) + 1]}$$

Equations E1 through E4 are the basis for A1 through A4. Cases with zero denominators are ignored. Non-zero denominators are always positive when $XRPB$, $BRRE$ and $ARRE$ are greater than one.

---

[37] If $XP=XQ=XF$ and $XN=XM=XH$, where $XF \equiv BH$ and $XH \equiv AH$, then $b_0=AQ+BQ\text{-}AF$, $b_1=AP\text{-}AQ$ and $b_2=BP\text{-}BQ$ *so* $E(0,0)=AQ+BQ\text{-}AF$, $E(0,1)=BP+AQ\text{-}AF$, $E(1,0)=AP+BQ\text{-}AF$ and $E(1,1)=AP+ BP\text{-}AF$. If $b_3 = 0$ then $Ra = E(0,0)$, $Rb = E(0,1)$, $Rc = E(1,0)$ and $Rd = E(1,1)$. If $AP=AQ$ the association is trivial and reversal is meaningless.

[38] $b_2(E|A,B)$ is obtained from $b_1(E|A,B)$ by exchanging A with B, AH with BH, AP with BP, XP with XN, and AF with AF. See D3b & D3c.

[39] If $b_1(E|A,B) = 0$, then $(AP\text{-}AQ)=(BP\text{-}BQ)(XP\text{-}XQ)$, so $b_2(E|A,B) = [(BP\text{-}BQ)\text{-}(BP\text{-}BQ)(XP\text{-}XQ)(XN\text{-}XM)]/D = (BP\text{-}BQ) = b_2(E|B)$.

[40] If $XP=XQ=XF$ and if $XN=XM=XH$, where $XF \equiv BH$ and $XH \equiv AH$, $E(0,0) = AF - [AH(AP\text{-}AQ)+BH(BP\text{-}BQ)]$, $E(0,1) = AF - [AH(AP\text{-}AQ)\text{-}(1\text{-}BH)(BP\text{-}BQ)]$. $E(1,0) = AF+[(1\text{-}AH)(AP\text{-}AQ)\text{-}BH(BP\text{-}BQ)]$, $E(1,1) = AF+[(1\text{-}AH)(AP\text{-}AQ)+(1\text{-}BH)(BP\text{-}BQ)]$. The form in footnote 37 is equivalent but simpler.