STATISTICAL LITERACY: SEEING HOW STATISTICAL SIGNIFICANCE IS CONTEXTUAL

Milo Schield, Augsburg College
Dept. of Business Administration. 2211 Riverside Drive. Minneapolis, MN 55454

Abstract: In observational studies, a statistically significant relationship can become statistically insignificant after taking into account the influence of a confounding factor. This is simply a result of the fact that in observational studies, all associations are contextual. Associations can change depending on what you take into account. First, a graph type is introduced that displays the influence of both binary predictor variables on a rate outcome. Secondly the non-interaction condition is introduced to simplify the presentation. Three cases are examined: an association is changed, does not disappear and is not reversed, an association is changed, does not disappear and is reversed, and an association is changed and disappears (is spurious). Third, confidence intervals are used to identify whether an observed difference between two percentages is statistically significant. The influence of a third factor is shown to be able to transform a statistically significant relation into one that is statistically insignificant. Although none of this is new, this approach to statistical significance and the influence of context is readily presented and easily understood. This approach is strongly recommended for those teaching introductory statistics.
Keywords: Epidemiology, Confidence Intervals

INTRODUCTION

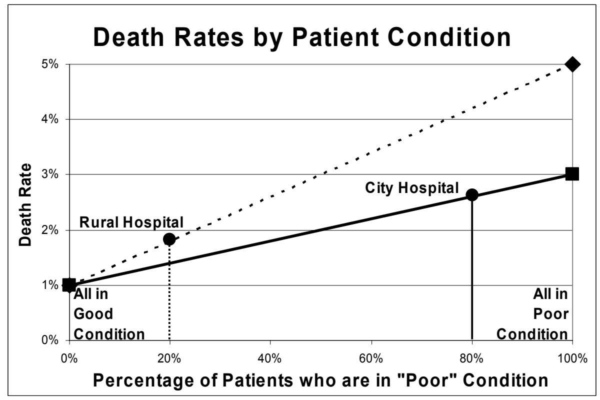
Statistical significance is a most important concept. But in most introductory statistics texts, the journey from the Binomial theorem through sampling distributions, the central mean, confidence intervals and hypothesis tests to statistical significance is a long journey often taking more than half of a semester course. Statisticians are keenly aware that in observational studies statistical significance depend critically on what is taken into account. But in an introductory course, there is little time to emphasize this fact.

This paper shows an approach that can be taught in the introductory course in a short amount of time. This paper presents student level teaching materials along with teacher level commentary. This approach combines the algebraic criteria for spuriousity (Schield and Burham, 2003), the graphical approach to displaying confounding (Wainer, 2002) and the short-cut approach to statistical significance (Giere, 1996; Burkholder and Giere, 2003). The goal is to help students understand – quickly and easily – what statistical significance is and how it can be contextual.

1. CONFOUNDING: GRAPHICAL APPROACH

A graphical approach to confounding is presented that applies to binary variables (Wainer, 2002). We will focus on one example involving death rates at two hospitals in relation to patient condition where patients are in either poor or fair condition.

Figure 1: Death rates by hospital and patient condition



Both hospitals have the same death rate for patients in good condition, but the rural hospital has a much higher death rate than the city hospital for those in poor condition. The overall death rate is just the weighted average of the two extremes depending on the mix of patients, so the mix-related pattern is always a straight line.

The rural hospital has mostly patients in good condition while the city hospital has mostly patients in poor condition, so the overall death rate at the rural hospital is lower than that at the city hospital. But once we take into account the condition of the patient, we see that for the same mix of patients, the rural hospital never has a lower death rate than the city hospital. This reversal is an instance of Simpson's Paradox. The relation between two weighted averages is the opposite of that between the components of the weighted averages.

2. NON-INTERACTION

Statisticians recognize there is no necessity for parallel lines. The presence of parallel lines simply reflects a lack of interaction between hospital and patient condition. The lack of interaction is readily seen using the following regression model,

$E = \text{constant} + aA + bB + cAB$.

For three binary variable it can be shown that constant = Ra, a = Rc-Ra, b = Rb-Ra, and c = (Rd-Rc)–(Rb-Ra). If the AB coefficient is zero, the (Rd-Rc) = (Rb-Ra), so the lines must be parallel. Thus, in Figure 1 the AB coefficient is non-zero, while in Figure 2 the AB coefficient is zero.

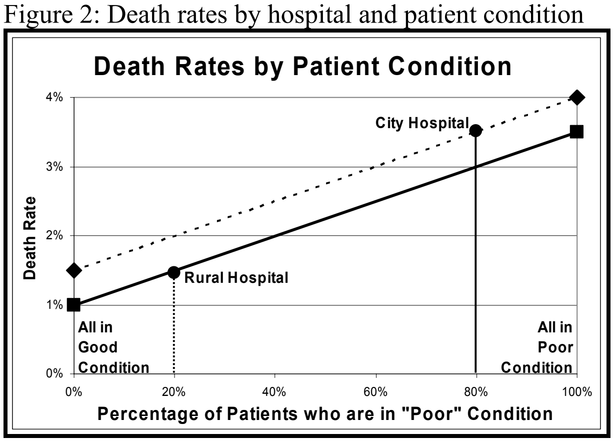
To simplify the presentation, non-interaction is stipulated so that the model of the data is readily obtained from the data.

3. CONFOUNDING: THREE CASES

Assuming non-interaction, we have three different cases in an observational study. For simplicity, the observed death rate in the city hospital is taken to be higher than that observed in the rural hospital.

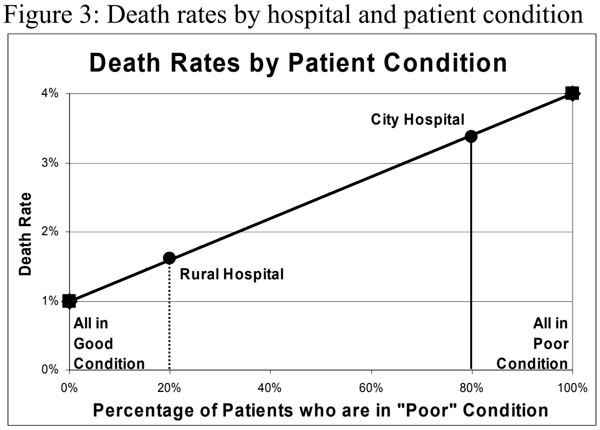
- In Figure 2, the rural hospital has a lower death rate than city hospital for the same mix of patients.
- In Figure 3, the rural hospital has the same death rate as the city hospital for the same mix of patients.
- In Figure 4, the rural hospital has a higher death rate than city hospital for the same mix of patients.

Figure 2: Death rates by hospital and patient condition



The observed death rate at the city hospital (3.5%) is greater than that observed at the rural hospital (1.5%). But when both hospitals are standardized to have the same mix of patients as in the population (50%), the difference in death rates shrinks from 2 percentage points (3.5% - 1.5%) to 0.5 percentage points. (Since the lines are parallel, the 2.75% - 2.25% = 4% - 3.5% = 1.5% - 1%.) Standardizing decreases the difference, but does not eliminate it or reverse it.

Figure 3: Death rates by hospital and patient condition

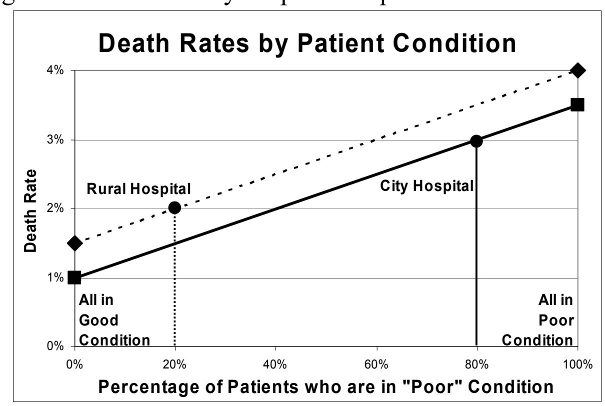


In Figure 3, the observed death rate at the city hospital (3.5%) is greater than that observed at the rural hospital (1.5%). But when both hospitals are standardized to have the same mix of patients as in the population the difference in death rates shrinks from 2 percentage points (3.5% - 1.5%) to zero. (Since the lines are parallel, the 2.5% - 2.5% = 4% - 4% = 1% - 1%.) Standardizing totally eliminates the difference.

In Figure 3, the observed association is completely spurious. Once we take into account the difference in patient condition, the observed association disappears completely and the standardized death rates are equal for the two hospitals.

If the coordinates of the two weighted averages are known (death rates and prevalences of confounder), then one can algebraically determine the necessary condition for a confounder to make the observed association completely spurious.

Figure 4: Death rates by hospital and patient condition



In Figure 4, the observed death rate at the city hospital (3%) is greater than that observed at the rural hospital (2%). But when both hospitals are standardized to have the same mix of patients as in the population, the difference in death rates reverses. It goes from having the city hospital's death rate 1 percentage point (3.5% - 1.5%) greater to having the rural hospital's death rate 0.5 percentage points greater. (Since the lines are parallel, the 2.75% - 2.25% = 4% - 3.5% = 1.5% - 1%.) Standardizing reverses the direction of the association. This is a clear instance of Simpson's paradox.

4. REDUCTION OF FACTORS

When there is no interaction between the confounder and the predictor, it seems that we have three factors at work: a difference in rates for the two values of the confounder (RD - RC) and difference in mixture between the treatment group and the population (XP-XF) and between the control group and the population (XF-XQ). If we designate that difference in rates as RK, then RK is non-zero in Figures 2 and 3, but zero in Figure 4.

If the difference in the prevalence of B is constrained by the prevalence of B in the population then we can reduce two of the variables to one. We will use Z to scale XP and XQ as follows:

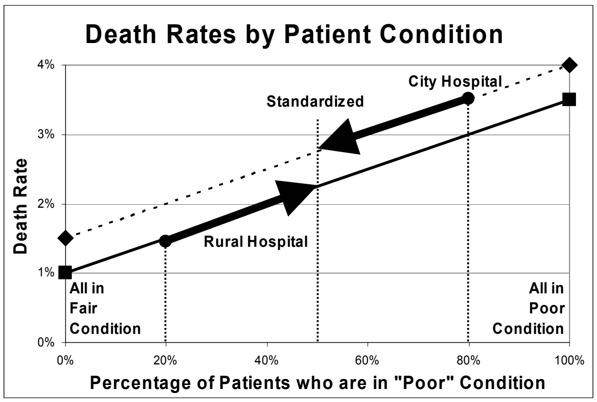
$$XP = XPo - Z(XPo - BH)$$
$$XQ = XQo + Z(BH - XQo)$$

This scaling maintains the observed prevalence of B regardless of the value of Z. If Z = 0, we get what was actually observed (without taking into account confounding). If Z = 1, we get what would be observed if we took into account confounding. Doing this reduces the number of factors to two: the difference in mixture (Z) and the difference in death rates.

5. STANDARDIZING

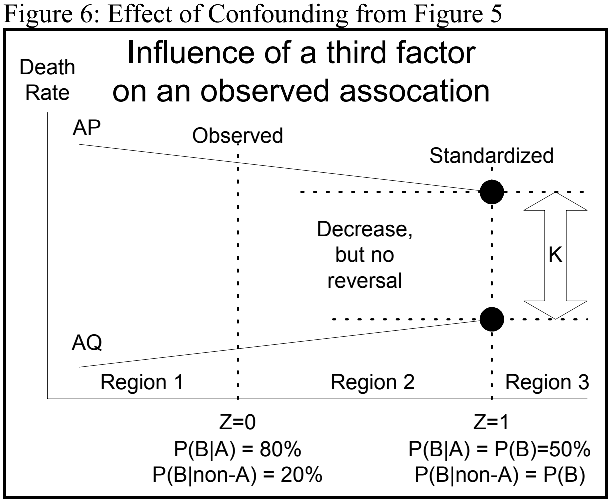
Standardizing gives each hospital the same mix of patients as is observed in the population. Consider first the situation shown in Figure 5.

Figure 5: Changing the Mix in Figure 2



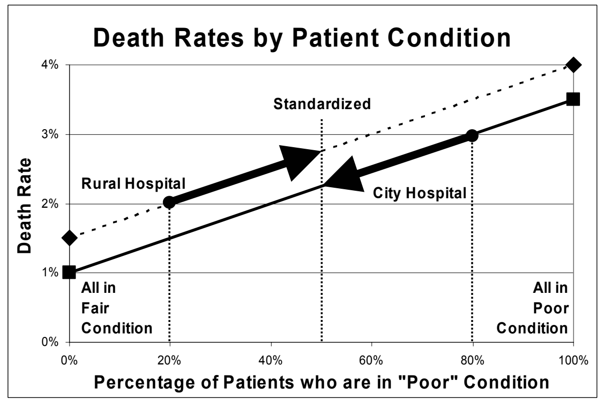
As we move the percentage of patients in poor condition from their observed values (Z = 0, P(B|A)=80%, P(B|non-A)=20%) to the percentage in the entire population (Z=1, P(B)=50%), the death rate at the city hospital decreases, while that at the rural hospital increases. Plotting the death rates for the two hospitals versus Z, we get the following graph for AP and AQ.

Figure 6: Effect of Confounding from Figure 5



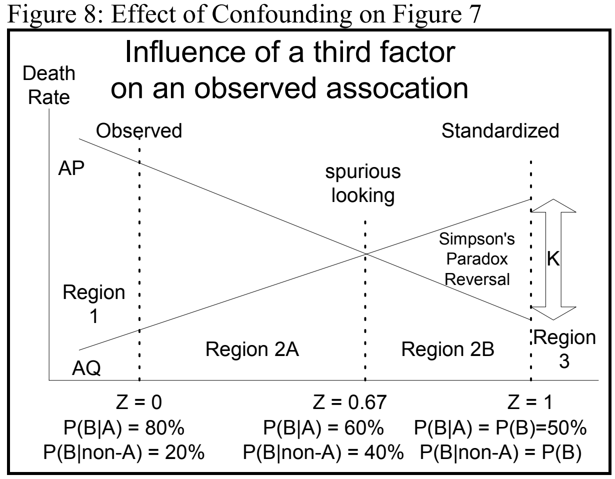
In Figure 6 we have two points and three regions. One point on the scale of influence is what is actually observed (Z=0). A second point is what would be observed after taking into account the influence of a confounder (Z=1). These two points give three regions. Region 2 is the region between what is actually observed and what would be observed after taking into account confounding. Region 1 is what would be observed if the mix were even more extreme than what was observed (Z < 0). Region 3 is what would be observed if the mix were reversed from what was observed (Z > 1).

Figure 7: Changing the Mix in Figure 4



As we move the percentage of patients in poor condition from their observed values (Z = 0, P(B|A)=80%, P(B|non-A)=20%) to the percentage in the entire population (Z=1, P(B) = 50%), the death rate at the city hospital decreases, while that at the rural hospital increases. But in this case, they pass a point at which they are equal and end up being reversed. The standardized death rate at the city hospital is lower than that at the rural hospital. Plotting the death rates for the two hospitals versus Z, we get the following graph for the values of AP and AQ.

Figure 8: Effect of Confounding on Figure 7



In Figure 8 we have the same two points as seen in Figure 6. The Z=0 point is what is actually observed. The Z=1 point is what would be observed after taking into account the influence of a confounder. Regions 1 and 3 are the same as before.

But in Figure 8, we have a new point: the point at which the association appears completely spurious and the association changes direction. This third point subdivides region 2 into two parts: 2A and 2B. In terms of the hospital example, the left edge of region 2 (Z=0) occurs when 80% of city hospital patients are in poor condition, while only 20% of rural hospital patients are in poor condition. The right edge of Region 2 (Z=1) occurs when 50% of the patients are each hospital are in poor condition. In this case, the cross-over point occurs when the percentage who are in poor condition is 60% for the city hospital and 40% for the rural hospital.

6. SAMPLING

Consider each of the observed fractions as samples from a larger population. If these are random samples, then we can construct confidence intervals for the observed fractions.

One may question whether confidence intervals are appropriate since it is not obvious that there is any sampling. Indeed, the hospital data may be the entire population for that hospital for a given time period. But our goal is to compare the overall death rates at the hospitals – not just for those in a particular time interval, but for those in general. As such, we can view what happened in a particular time interval as a sample taken from the larger population of death rates throughout time.

This is no different that what is done in clinical trials where the data represents everybody in the study, yet we still do hypothesis tests to see that the observed difference is statistically significant so we can say something about the difference in the larger populations of those who might have received the treatment and those who might not have.

Although this may seem to have a Bayesian flavor, it does not involve taking something that is constant in nature (e.g., the speed of light) and talking about the probability that that value is in a particular confidence interval. This is like a medical test where different samples can have different percentages of subjects who have the disease. This form of sampling occurs whenever we have a generation process generating outcomes sequentially.

We can attach confidence intervals to any pattern of observations whether it is a straight line or not. But if the pattern is a straight line (see figure 9) then the bands formed by the confidence intervals will involve straight lines as well.

7. STATISTICAL SIGNIFICANCE

If the confidence intervals overlap, then we definitely have a lack of statistical significance. Now consider the results of attaching confidence intervals to the proportions shown in Figure 9.

Figure 9: Confounding and statistical significance

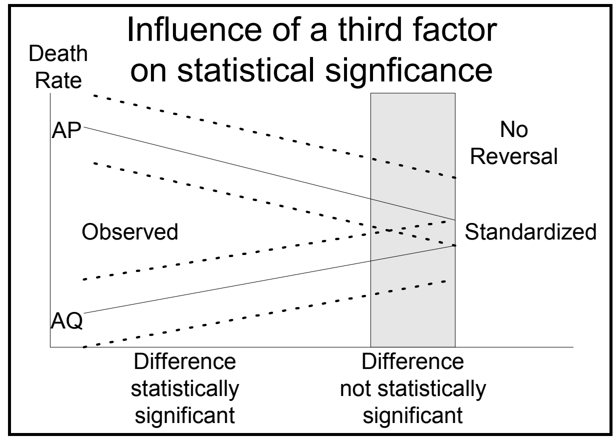
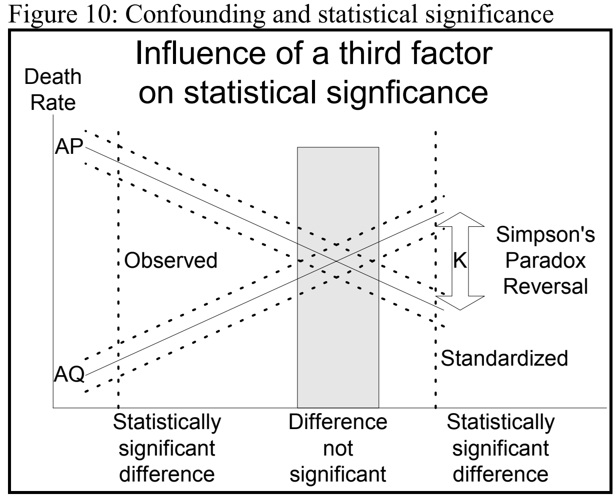


Figure 10: Confounding and statistical significance



The region in which the difference is not statistically significant is centered on the point at which the association is 100% spurious.

There is no requirement that the third factor be sufficient to bring about a Simpson's paradox reversal. The influence may simply decrease the observed association as shown in Figure 9.

8. DISCUSSION

In a well-designed experiment, subjects are randomly assigned to the treatment and control groups. If the resulting difference is determined to be statistically significant, that conclusion is treated as being absolute – it is not expected to be contextual. The influence of any plausible confounder is expected to be zero due to the random assignment involved in assigning subjects to the experiment. So if an association in a well-designed experiment is statistically significant, one can conclude that the treatment caused the observed association or the difference.

But in an observational study, there are an unlimited number of plausible confounders that might be taken into account. And when an observed association is found to be statistically significant, that conclusion is heavily contextual. So finding that an association is statistically significant in an observational study is weak evidence for presuming that a standardized association would be statistically significant much less for presuming that the difference between exposure and non-exposure caused the observed difference in outcomes.

Understanding the difference between an experiment and an observational study is critical in understanding whether an association is contextual and whether being statistically significant is contextual.

9. CONCLUSION

In summary, we can see that in observational studies, statistical significance is contextual. This follows from the fact that in observational studies, associations are contextual. Recognizing that an association can be contextual is a fundamental goal of being a critical thinker. Recognizing that statistical significance can be contextual is extremely important in understanding the role of statistics as evidence.

Taking into account the influence of a confounder can transform an association that is statistically significant into one that is not. Recognizing this possibility – indeed this all too often condition – is fundamental to thinking critically about any claim based on data from an observational study.

Students studying statistical literacy must understand the nature and importance of statistical significance, and they must understand when it is absolute (well-designed experiments) and when it is contextual (all observational studies).

10. REFERENCES

Burkholder, Leslie and Ronald Giere (2003). *Understanding Scientific Reasoning.* 2003 ASA Proceedings of the Section on Statistical Education.

Giere, Ronald N. (1996). *Understanding Scientific Reasoning.* 4th edition, Holt, Rinehart and Winston.

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

Schield, Milo and Thomas Burnham (2003). *Algebraic Conditions for a Spurious Association Between Two Binary Variables.* 2003 ASA Proceedings of the Section on Statistical Education.

Wainer, Howard (2002). "*The BK-Plot: Making Simpson's Paradox Clear to the Masses*" Chance Magazine Vol. 15, No. 3, Summer 2002, pp. 60–62.

Acknowledgments: This research was supported by a grant from the W. M. Keck Foundation to Augsburg College "to support the development of statistical literacy as an interdisciplinary discipline

Contact: Dr. Schield is at Schield@Augsburg.edu.
This paper is at www.augsburg.edu/ppages/~schield.

APPENDIX A: ALGEBRAIC RELATIONS BETWEEN THREE BINARY VARIABLES

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***	***Bn = 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***	***Xn = 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*	*RG=Ab/Ag*
A	*RC=Ec/Xc*	*RD=**Ed/Xd***	*RH=**Ad/Ah***
TOTAL	*RE=Bb/Bg*	*RF=**Bd/Bh***	*RN = **Af/n***

E designates an effect; A and B designate predictors. In names, the first letter designates the table. Tables A, B, X and E are count tables (R is a ratio table).

DEFINITIONS:

a. AP = Risk of E for A = Ad/Ah
b. AQ = Risk of E for non-A = Ab/Ag
c. XP = Prevalence of B for A = $Xd/Xh = AH$
d. XQ = Prevalence of B for non-A = $Xb/Xg = AG$
e. BH = Prevalence of B = Bh/n

Margin Value Ratios:
$AP = RH$, $AQ = RG$. $BP = RF$, $BQ = RE$
f. $RH = (RD\cdot XD + RC\cdot XC)/XH = RD\cdot XP + RC(1\text{-}XP)$
g. $RG = (RB\cdot XB + RA\cdot XA)/XG = RB\cdot XQ + RA(1\text{-}XQ)$
h. $RF = (RD\cdot XD + RB\cdot XB)/XF = RD\cdot XN + RB(1\text{-}XN)$
i. $RE = (RC\cdot XC + RA\cdot XA)/XE = RC\cdot XM + RA(1\text{-}XM)$

NON-INTERACTION BETWEEN A AND B.

1a. $RD - RB = RC - RA$ Non-interaction criteria
1b. $RD - RC = RB - RA$ Equivalent form
1c. $RA + RD = RB + RC$ Equivalent form

If $XP = XQ$, then [1]
2a. $RH - RG = RD - RB = RC - RA$
2b. $RF - RE = RD - RC = RB - RA$

3a. Let $RJ = RD - RB = RC - RA$,
so $RD = RB + RJ$ and $RC = RA + RJ$.
3b. Let $RK = RD - RC = RB - RA$
so $RD = RC + RK$ and $RB = RA + RK$.
3c. $RJ - RK = RC - RB$

4a $AP = RH = RD\cdot XP + RC(1\text{-}XP)$
4b $AQ = RG = RB\cdot XQ + RA(1\text{-}XQ)$
4c $BP = RF = RD\cdot XN + RB(1\text{-}XN)$
4d $BQ = RE = RC\cdot XM + RA(1\text{-}XM)$

5a $AP = (RC\cdot + RK)XP + RC\cdot(1 - XP)$
5b $AQ = (RA\cdot + RK)XQ + RA\cdot(1 - XQ)$
5c $BP = (RB\cdot + RJ)XN + RB(1 - XN)$
5d $BQ = (RA\cdot + RJ)XM + RA(1 - XM)$

6a $AP = RC\cdot + RK\cdot XP$. $AQ = RA\cdot + RK\cdot XQ$[2]
6c $BP = RB\cdot + RJ\cdot XP$. $BQ = RA\cdot + RJ\cdot XP$

7a. $AP - AQ = (RC\cdot + RK\cdot XP) - (RA\cdot + RK\cdot XQ)$
7b. $AP - AQ = RJ + RK(XP - XQ)$[3]

8a. Let $XP = XPo - Z(XPo - BH)$
8b. Let $XQ = XQo + Z(BH - XQo)$
8c. $XP\text{-}XQ = (XPo\text{-}Xqo) - Z[(XPo\text{–}BH)+(BH\text{-}XQo)]$
8d. $XP\text{-}XQ = (XPo\text{-}Xqo)(1 - Z)$[4]

9a. $AP - AQ = RJ + RK(XPo - XQo)(1\text{-}Z)$
9b. If $AP = AQ$, $(1\text{-}Z) = \text{-}RJ/[RK(XPo - XQo)]$[5]

---

[1] 2c. $RH\text{-}RG=[RD\cdot XP+RC(1\text{-}XP)]-[RB\cdot XQ+RA(1\text{-}XQ)]$
2d. $RH\text{-}RG=[RD\cdot XP+RC(1\text{-}XP)]-[RB\cdot XP+RA(1\text{-}XP)]$
2e. $RH\text{-}RG=(RD\text{-}RB)XP + (RC\text{-}RA)(1\text{-}XP)$
2f. $RH\text{-}RG=(RD\text{-}RB)XP + (RD\text{-}RB)(1\text{-}XP)$
2g. $RH\text{-}RG = RD\text{-}RB$. QED.

[2] Let XP=.8 and XQ=.2.
If RC = .010, RK=.025, then AP = .03. OK
If RA = .015, RK = .025, then AQ = .02. OK

[3] Check: RC-RA = (.010 - .015) = -.005.
RK*(XP – XQ) = .025(0.8 - .2) = .015
AP-AQ = -.005 + .015 = +.010. OK

[4] If XPo = .8, XQo = .2,
If Z = 1, XP=XQ. OK
If Z = 1/3, XP-XQ=.4 OK
IF Z = 0, XP-XQ = XPo – XQo OK
9a AP = RC·+ RK·[XPo – Z(XPo – BH)]
9b AQ = RA·+ RJ·[XQo + Z(BH – XQo)]

[5] (1-Z) = -(-.005)/[.025(.8-.2)] = .005/.015 = 1/3. NO!