

08/10/2002 1

Epidemiology and Statistics

Algebraic Associations in 2x2 Tables

MILO SCHIELD
*Director, W. M. Keck Statistical Literacy Project
 Augsburg College, Minneapolis MN*

THOMAS V.V. BURNHAM
Cognitive Consulting, San Antonio TX

www.augsburg.edu/ppages/~schiel schield@augsburg.edu

08/10/2002 2

Epidemiology and Statistics: SIMILARITIES

Both share

- a common foundation in probability
- a common focus on chance
- a common respect for confounding
- a common interest in modeling
- a common basis in empirical data

Yet epidemiology and statistics are different!

08/10/2002 3

Epidemiology and Statistics: DIFFERENCES

	Epidemiology	Statistics
Samples	non-representative	representative
Data	Observational	Experimental
Association	RR and OR	Phi
Goal	Prevent bad outcomes	Predict outcomes of interest

RR: Relative Risk; **OR:** Odds Ratio
Phi: Pearson correlation coefficient for binary data

08/10/2002 4

Epidemiology and Statistics: STRENGTH OF ASSOCIATION

	Epidemiology	Statistics
Strong	RR > 10	Phi > 0.9
Weak	RR < 1.2	Phi < 0.1
Smoking & lung cancer	RR = 11	Phi = 0.06

**Is this association strong or weak?
 Epidemiologists use RR; Phi isn't "useful"...**

08/10/2002 5

Epidemiology and Statistics: WHY ISN'T PHI USEFUL?

Phi

- requires a representative sample.
- measures sufficiency as well as necessity.
- measures relative to the perfect predictor: 100% necessary and 100% sufficient.

$$Phi^2 = N \cdot S[(RR-1)/RR] / [(RP-1)/RP]$$

$$Phi^2 = (N-H)(S-F) / [(1-H)(1-F)]$$

08/10/2002 6

Epidemiology and Statistics: RELATIVE RISK & RELATIVE PHI

Phi: Pearson correlation coefficient for binary data
F: Risk of case (E.g., 68 lung cancers /1,000 deaths)
H: Prevalence of predictor (E.g., 30% for smokers)

$$Phi^2 = [F/(1-F)][H/(1-H)] \{ [H(RR-1)] / [H(RR-1) + 1] \}^2$$

RPhi: Phi *relative to* the Phi for a 100% necessary predictor having the same prevalence

$$RPhi^2 = Phi^2(RR, H, F) / Phi^2(RR=\infty, H, F)$$

$$RPhi^2 = \{ [H(RR-1)] / [H(RR-1) + 1] \}^2$$

Smoking and lung cancer: RPhi = .7 if RR = 11.

08/10/2002 7

Epidemiology: NECESSITY & SUFFICIENCY

Necessity and sufficiency are determinate (binary).
Consider treating them as probabilistic (fractions).

- Sufficiency: % of smokers who develop lung cancer
- Necessity: % of lung cancer patients who smoked

Epidemiology is more concerned with

- eliminating disease than identifying causes.
- with necessity than sufficiency.

Relative risk indicates necessity, *not* sufficiency

$$RR = [N/(1-N)] / [H/(1-H)]$$

08/10/2002 8

Epidemiology: NECESSITY AND CAUSALITY

Suppose we treat any factor that is “more necessary than not” ($N > 0.5$) as “causal.”

$$RR = [N/(1-N)] / [H/(1-H)].$$

If $N > 0.5$, then $RR > (1-H)/H$

$$[\text{If } H = 0.25, \text{ then } RR > 3]$$

Examples with high degree of necessity ($H=.25$):

- Well water *caused* cholera ($RR=14$; $N=.82$)
- Smoking *causes* lung cancer ($RR=10$; $N=.77$)

08/10/2002 9

Epidemiology: BAYES RATIO

Bayes Ratio is commonly used in the news:

Unmarried pregnant women are 80% more likely to abort than are *pregnant women*.

Pregnant women who abort are 80% more likely to be *unmarried* than *pregnant women*.

Bayes rule allows conversion of compares.

08/10/2002 10

Epidemiology: BAYES RATIO to RELATIVE RISK

Unmarried: 45% of pregnant women; 81% of abortions.

Married: 55% of pregnant women; 19% of abortions.

Unmarried pregnant women are 1.8 times as likely to have an abortion as are all pregnant women.

Married pregnant women are 0.35 times as likely to have an abortion as are all pregnant women.

Among pregnant women, unmarried women are 5.2 times as likely to have an abortion as are married women.

08/10/2002 11

The Epidemiological Approach: INCREASINGLY IMPORTANT

- is data-based and decision/action oriented.
- is quantitatively-stated and scientifically supported.
- is used increasingly in public health (e.g., deaths attributable to pollution from a power-plant).
- is spreading to include social situations (e.g., gun ownership, watching violence on TV).
- is becoming a basis for public policy (e.g., ban on second hand smoke with $RR = 1.19$).

08/10/2002 12

Statistics Education: RECOMMENDATIONS

Statistics education should include more on

- the epidemiological approach.
- ratios and comparisons of ratios.
- Bayes rule and Bayes comparison.
- probabilistic causation.
- probabilistic necessity and sufficiency.
- causation inferred from observational data.
- relative risk criteria for non-confounding.

08/10/2002 13

Epidemiology: OVER-INVOLVEMENT

Over-involvement (an ordinal association) means “more likely to be involved than is expected in the population.”

A. If Target shoppers are more likely to clip coupons than are the non-buyers, then Target buyers are *over-involved* in clipping coupons.

B. If coupon clippers are more likely to shop at Target than are non-coupon clippers, then coupon clippers are *over-involved* in shopping at Target.

If A is true, then B is true (and vice-versa).

08/10/2002 14



Milo Schield

Tom Burnham