17: Odds Ratios

Independent Samples | Matched-Pairs

Independent Samples

Introduction

The prior chapter used risk ratios from cohort studies to quantify exposure–disease relations. Briefly, cohort studies follow groups of people in a closed population (cohort) to determine and compare incidences. One of the limiting factors of cohort studies is they require the study of many individuals over extended periods of time to accumulate sufficient cases for study, especially when the disease is rare. Case-control studies overcome this limitation by studying all cases from the population but only a subset of the noncases.

Because case-control studies lack information about the size of the population at risk, they are unable to estimate risks directly. This precludes use of risk differences and risk ratios as measure of association. Thus, a different statistic is needed. This need is fulfilled by odds ratio. (The sampling basis of the odds ratio is discussed on pp. 208–212 in *Epidemiology Kept Simple*.)

As before, data are cross-tabulated as follows:

	Disease+	**Disease−**	
Exposure +	A_1	B_1	N_1
Exposure -	A_0	B_0	N_0
	M_1	M_0	N

The proportion of case who are exposed ($p_1 = A_1 / M_1$), and proportion of control who are exposed ($p_0 = B / M_0$). However, the more interesting comparison comes in the form of the **odds ratio,** denoted *OR* or ψ (lower case psi). The **odds ratio estimator** is:

$$\hat{y} = \frac{A_1 / A_0}{B_1 / B_0} = \frac{A_1 B_0}{B_1 A_0} \quad \textbf{(1)}$$

which is merely the cross-product ratio from the table. This statistic estimates either the rate ratio or risk ratio in the source population depending on how controls were selected for study. Either way, it is a measure of relative incidence of disease conditional on exposure.

Illustrative example. Breslow and Day (1980) report data from a case-control study of 200 esophageal cancer cases and 775 community-based controls (Tuyns, 1977). A detailed dietary questionnaire about alcohol consumption, tobacco use, and other factors was administered to study participants. Cross-tabulation of alcohol consumption (dichotomized at 80 grams per day) and esophageal cancer reveals:

Alcohol (g/day)	Cases	Controls	
≥ 80	96	109	205
< 80	104	666	770
	200	775	975

Thus, $\hat{y}$ = (96)(666)/(109)(104) = 5.64. This suggests esophageal cancer was 5.64 times as frequent in the exposed group than in the nonexposed groups in the source population.

95% Confidence Interval for the ψ: Let ψ denote the odds ratio parameter. To calculate a 95% confidence interval for ψ, convert the odds ratio estimate ($\hat{y}$) to a natural logarithmic scale and then calculate its standard error as $se_{\ln \hat{y}} = \sqrt{\frac{1}{A_1} + \frac{1}{A_0} + \frac{1}{B_1} + \frac{1}{B_0}}$. The 95% confidence interval for the lnψ is:

$$\ln \hat{y} \pm (1.96)(se_{\ln \hat{y}}) \qquad \textbf{(2)}$$

Anti-logs are taken to derive the confidence limits for the ψ.

Illustrative example. For the illustrative data, ln(5.64) = 1.7299, $se_{\ln \hat{y}} = \sqrt{\frac{1}{96} + \frac{1}{109} + \frac{1}{104} + \frac{1}{666}}$ = 0.1752, and the 95% confidence interval for lnψ = 1.7299 ± (1.96)(0.1752) = 1.7299 ± 0.343392 = (1.387, 2.073). Thus, the 95% confidence interval for ψ = $e^{1.387,\ 2.073}$ = (4.00, 7.95).

90% Confidence Interval. The level of confidence for the interval can be altered by changing the 1.96 in the above equation to 1.645. Thus, a 90% confidence interval for the lnψ of the illustrative data is 1.7299 ± (1.645)(0.1752) = 1.7299 ± 0.2882 = (1.4417, 2.0181) and a 90% confidence interval for the ψ = $e^{(1.4417,\ 2.0181)}$ = (4.23, 7.52).

p Value. A p value for H_0: ψ = 1 is calculated with a chi-square statistic or Fisher's test, depending on circumstances (see prior chapter). For the illustrative example, χ^2_{Yates} = 108.22 with 1 degree of freedom ($p \approx$ .000001).

Computations may be assisted with WinPepi or EpiCalc2000.

Odds Ratios Based on Matched-Pairs

Introduction. A matched design may be used to help control for potential confounders. When matching is employed, a different method of analysis is required. Matched-pair data are displayed as follows:

	Control pair-member exposed	Control pair-member nonexposed
Case pair-member exposed	*t*	*u*
Case pair-member nonexposed	*v*	*w*

Cells *t* and *w* in this table contain counts of **concordant pairs** (pair members are the same with respect to exposure), while cells *u* and *v* contain **discordant pairs** (pair members differ with respect to exposure).

Estimation. The **odds ratio estimator** for matched data is:

$$\hat{y} = \frac{u}{v} \tag{3}$$

The **standard error of this natural log of the estimator** is $se_{\ln \hat{y}} = \sqrt{\frac{1}{u} + \frac{1}{v}}$, and the **95% confidence interval for the ln**ψ is

$$\ln \hat{y} \pm (1.96)(se_{\ln \hat{y}}) \tag{4}$$

Anti-logs provide confidence limits on the non-logarithmic scale.

***Illustrative example*.** Fifty matched-pairs (100 people) are distributed as follows:

	Control pair-member exposed	Control pair-member is nonexposed
Case pair-member is exposed	5	30
Case pair-member nonexposed	10	5

Thus,

- $\hat{y}$ = 30 / 10 = 3.000
- $se_{\ln\psi}$ = %(1/30 + 1/10) = 0.3651
- 95% confidence interval for lnψ = ln(3.000) ± (1.96)(0.3651) = 1.0986 ± .7156 = (0.3830, 1.8142)
- 95% confidence interval for ψ = $e^{(0.3830, 1.8142)}$ = (1.47, 6.14).

The odds ratio estimate and confidence interval are interpreted in the usual manner (i.e., point estimate: exposure triples risk; we are 95% confident the odds ratio parameter lies between 1.47 and 6.14.

McNemar's Test. The usual chi-square and *z* tests do not apply to matched-pair data. Instead, use McNemar's chi-square statistic. The continuity corrected McNemar's chi-square is as follows:

$$\chi^2_{McN,corrected} = \frac{(|u - v| - 1)^2}{u + v} \tag{5}$$

This statistic has 1 degree of freedom.

The statistic may be converted to a chi statistic by taking its square root, thus facilitating computation of a *p* value.

The $\chi^2_{\text{McN,Cor.}} = (|30 - 10| - 1)^2 / (30 + 10) = 9.025$, $p = .0027$.