

Introduction to Mantel-Haenszel estimate

- The two examples in lecture3 have illustrated the reason why it is not appropriate to use marginal (crude) odds ratio to examine the association of the exposure variable (x) with the response variable (y), and the need to use conditional (adjusted) odds ratios. Therefore, the population parameters of interest are those conditional (adjusted) odds ratios rather than marginal (crude) odds ratio.

Introduction to Mantel-Haenszel estimate

- Q: Note that the data in both examples in lecture 3 are population data which allows one to know exactly what those conditional (adjusted) odds ratios are. In reality, one only has sample data. As before the natural question to ask is how to estimate those conditional (adjusted) odds ratios based on sample data?

Introduction to Mantel-Haenszel estimate

- The answer to the question depends on whether or not those conditional (adjusted) odds ratios are different across the levels of Z :
 - Case 1: those conditional (adjusted) odds ratios are different. This case is the case where we call z as **effect modifier**.
 - Case 2: conditional (adjusted) odds ratios are the same

Introduction to Mantel-Haenszel estimate

- In case 1, the conditional (adjusted) odds ratio that corresponds to a specific level of z can be estimated using the 2 by 2 table of y and x that corresponds to that level of z .

Introduction to Mantel-Haenszel estimate

- In case 2, as in both examples in lecture 3, the common conditional (adjusted) odds ratio can be estimated by so-called **Mantel-Haenszel estimate of common odds ratio**.

Introduction to Mantel-Haenszel estimate

- Let θ denote the common conditional (adjusted) odds ratio. Obviously, an Ad-hoc approach to estimating θ is to use one of the z-level-specific 2 by 2 tables of y and x.
- The disadvantage of this Ad-hoc approach is the fact that it fails to use all the data of y and x, which leads to lose of efficiency, i.e., wider confidence interval.

Introduction to Mantel-Haenszel estimate

- Q: how to combine 2 by 2 tables together to estimate θ ?
- A: Three types of estimates of θ :
 - maximum likelihood estimate
 - Mantel-Haenszel estimate
 - Logit Estimate

Likelihood estimate of common odds ratio

- The likelihood estimate of θ won't be covered here since it will be discussed in the logistic regression context.

Mantel-Haenszel estimate of common odds ratio

- Mantel and Haenszel(1959) proposed a computationally simpler estimate for the common conditional (adjusted) odds ratio, which is called *Mantel-Haenszel estimate of common odds ratio*
- To express this estimate, we need some notations

Mantel-Haenszel estimate of common odds ratio

- The individual data on (y,x,z) can be represented by r two by two tables of y and x with the k -th table corresponding to the k -th level of z . The k -th table is denoted as follows:

$x \backslash y$	1	0	
1	n_{k11}	n_{k10}	n_{k1}
0	n_{k01}	n_{k00}	n_{k0}
	m_{k1}	m_{k0}	n_k

Mantel-Haenszel estimate of common odds ratio

- *Mantel-Haenszel estimate of common odds ratio* takes the form:

$$\hat{\theta}_{MH} = \frac{\sum_{k=1}^r n_{k11}n_{k00} / n_k}{\sum_{k=1}^r n_{k10}n_{k01} / n_k}$$

Mantel-Haenszel estimate of common odds ratio

- *Mantel-Haenszel estimate of common odds ratio* can be viewed as a weighted average of the individual odds ratios:

$$\hat{\theta}_{MH} = \sum_{k=1}^r w_k \hat{\theta}_k$$

where

$$w_k = \frac{n_{k10}n_{k00}/n_k}{\sum_{j=1}^r n_{j10}n_{j00}/n_j}$$

is the weight associated with the k-th odds ratio estimate, $\hat{\theta}_k = n_{k11}n_{k00}/n_{k10}n_{k01}$. The weight w_k approximate the inverse variance of $\hat{\theta}_k$ when θ is near 1.

Mantel-Haenszel estimate of common odds ratio

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58    4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72    6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11    7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21    8 0 0 36
;
run;
proc freq data=cmh order=data;
  weight count;
  table center*smoke*cancer/ cmh ;
run;
```

Note

- The CMH option requests the Mantel-Haenszel and logit estimates of the odds ratios and the corresponding confidence intervals, as well as the p-values for both Breslow-Day and Cochran-Mantel-Haenszel tests .

Mantel-Haenszel estimate of common odds ratio

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.1745	1.9840	2.3832
	Logit	2.1734	1.9829	2.3823
Cohort (Col1 Risk)	Mantel-Haenszel	1.5192	1.4417	1.6008
	Logit	1.5132	1.4362	1.5942
Cohort (Col2 Risk)	Mantel-Haenszel	0.6999	0.6721	0.7290
	Logit	0.7011	0.6734	0.7300

Asymptotic Confidence Interval for common odds ratio

- Fact: When n_k are sufficiently large,
 $\log(\hat{\theta}_{MH})$ is approximately normally
distributed with mean and standard
error:

$$E(\log(\hat{\theta}_{MH})) = \log(\theta)$$

$$\sigma(\log(\hat{\theta}_{MH})) \text{ (too complicated)}$$

Asymptotic Confidence Interval for common odds ratio

- The asymptotic confidence interval for the log of θ is

$$\log(\hat{\theta}_{MH}) \pm z_{1-\alpha/2} \sigma(\log(\hat{\theta}_{MH}))$$

- The asymptotic confidence interval for θ can be obtained by exponentiating endpoints of the above confidence interval.

Asymptotic Confidence Interval for common odds ratio

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58     4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72     6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11     7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21     8 0 0 36
;
run;
proc freq data=cmh order=data;
  weight count;
  table center*smoke*cancer/ cmh ;
run;
```

Asymptotic Confidence Interval for common odds ratio

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.1745	1.9840	2.3832
	Logit	2.1734	1.9829	2.3823
Cohort (Col1 Risk)	Mantel-Haenszel	1.5192	1.4417	1.6008
	Logit	1.5132	1.4362	1.5942
Cohort (Col2 Risk)	Mantel-Haenszel	0.6999	0.6721	0.7290
	Logit	0.7011	0.6734	0.7300

Logit Estimate of Common Odds Ratio

- An alternative estimate of common odds ratio is called *logit estimate of common odds ratio*. The idea of this estimate is to estimate $\mu = \log(\theta)$ first, and then exponentiate it to get the estimate of θ .

Logit Estimate of Common Odds Ratio

- Calculating this estimate takes two steps:

Step 1: Estimate $\mu = \log(\theta)$ by a weighted average of the individual log odds ratios

$$\hat{\mu} = \sum_{k=1}^r w_k \log(\hat{\theta}_k)$$

where

$$w_k = \frac{\sigma_k^2}{\sum_{j=1}^r \sigma_j^2}$$

$$\sigma_k^2 = \left\{ \frac{1}{n_{k11}} + \frac{1}{n_{k10}} + \frac{1}{n_{k01}} + \frac{1}{n_{k00}} \right\}^{-1}$$

$$\log(\hat{\theta}_k) = \log(n_{k11}n_{k00}/n_{k10}n_{k01})$$

Logit Estimate of Common Odds Ratio

- Step 2: the estimate of θ is obtained by exponentiating the estimate of $\mu = \log(\theta)$, i.e.,

$$\hat{\theta}_L = \exp(\hat{\mu})$$

- The logit estimate is also reasonable estimate of common odds ratio, but it has problem with zero cell as opposed to M-H estimate.

Logit Estimate of Common Odds Ratio

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58    4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72    6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11    7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21    8 0 0 36
;
run;
proc freq data=cmh order=data;
  weight count;
  table center*smoke*cancer/ cmh ;
run;
```

Logit Estimate of Common Odds Ratio

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.1745	1.9840	2.3832
	Logit	2.1734	1.9829	2.3823
Cohort (Col1 Risk)	Mantel-Haenszel	1.5192	1.4417	1.6008
	Logit	1.5132	1.4362	1.5942
Cohort (Col2 Risk)	Mantel-Haenszel	0.6999	0.6721	0.7290
	Logit	0.7011	0.6734	0.7300

Asymptotic Confidence Interval for common odds ratio

- Fact: When n_k are sufficiently large, $\log(\hat{\theta}_L)$ is approximately normally distributed with mean and standard error:

$$E(\log(\hat{\theta}_L)) = \log(\theta)$$

$$\sigma(\log(\hat{\theta}_L)) = \left[\sum_{k=1}^r \sigma_k^2 \right]^{-1/2}$$

Asymptotic Confidence Interval for common odds ratio

- The asymptotic confidence interval for the log of θ is

$$\log(\hat{\theta}_L) \pm z_{1-\alpha/2} \sigma(\log(\hat{\theta}_L))$$

- The asymptotic confidence interval for θ can be obtained by exponentiating endpoints of the above confidence interval.

Asymptotic Confidence Interval for common odds ratio

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58     4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72     6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11     7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21     8 0 0 36
;
run;
proc freq data=cmh order=data;
weight count;
table center*smoke*cancer/ cmh ;
run;
```

Asymptotic Confidence Interval for common odds ratio

Estimates of the Common Relative Risk (Row1/Row2)

Type of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	2.1745	1.9840	2.3832
	Logit	2.1734	1.9829	2.3823
Cohort (Col1 Risk)	Mantel-Haenszel	1.5192	1.4417	1.6008
	Logit	1.5132	1.4362	1.5942
Cohort (Col2 Risk)	Mantel-Haenszel	0.6999	0.6721	0.7290
	Logit	0.7011	0.6734	0.7300

Breslow-Day Test

- The Mantel-Haenszel(logit) estimate of common odds ratio are developed under the hypothesis that the conditional odds ratios are equal. It is necessary to test this odds ratio homogeneity hypothesis:

$$H_0 : \theta_1 = \dots = \theta_r$$

before obtaining the Mantel-Haenszel(logit) estimate, where θ_k is the conditional odds ratio corresponding to the k-th level of z (k=1,...,r)

Non-central Hypergeometric distribution

- Fact: For the k th 2 by 2 table, the conditional distribution of n_{k11} given column totals, m_{k1} and m_{k0} , and row totals, n_{k1} and n_{k0} , fixed is so-called *Non-central Hypergeometric distribution*, which has the following probability mass function:

$$P(n_{k11} = i) = \frac{\binom{n_{k1}}{i} \binom{n_{k0}}{m_{k1}-i} \theta_k^i}{\sum_u \binom{n_{k1}}{u} \binom{n_{k0}}{m_{k1}-u} \theta_k^u}$$

Non-central Hypergeometric distribution

- Hypergeometric distribution is a special case of Non-central Hypergeometric distribution since when the odds ratio, θ_k , equals 1, the mass function of Non central Hypergeometric distribution becomes that of Hypergeometric distribution:

$$P(n_{k11} = i) = \frac{\binom{n_{k1}}{i} \binom{n_{k0}}{m_{k1}-i}}{\sum_u \binom{n_{k1}}{u} \binom{n_{k0}}{m_{k1}-u}} = \frac{\binom{n_{k1}}{i} \binom{n_{k0}}{m_{k1}-i}}{\binom{n_k}{m_{k1}}}$$

Breslow-Day Test

- The idea of Breslow-Day test is under the null hypothesis(θ_k are equal), n_{k11} is approximately Non-central Hypergeometric distributed with

$$\theta_k = \hat{\theta}_{MH}$$

and hence n_{k11} should be close to $E(n_{11k}; \hat{\theta}_{MH})$, the mean of this Non-central Hypergeometric distribution.

Breslow-Day Test

- Breslow-Day test statistics takes the form:

$$\chi_{BD}^2 = \sum_{k=1}^r \frac{[n_{k11} - E(n_{k11}; \hat{\theta}_{MH})]^2}{Var(n_{k11}; \hat{\theta}_{MH})}$$

- Under H_0 , Breslow-Day test statistics has a chi-squared distribution with degrees of freedom $r-1$.

Breslow-Day Test

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58    4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72    6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11    7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21    8 0 0 36
;
run;
proc freq data=cmh order=data;
  weight count;
  table center*smoke*cancer/ cmh ;
run;
```

Breslow-Day Test

Breslow-Day Test for
Homogeneity of the Odds Ratios

Chi-Square	5.199
DF	7
Pr > ChiSq	0.6356

Cochran- Mantel-Haenszel Test

- Cochran- Mantel-Haenszel test is to test whether the common conditional (adjusted) odds ratio of y and x equals to one, i.e.

$$H_0 : \theta = 1$$

- Of course, one can use the confidence interval of θ to test this null hypothesis. The problem with using confidence interval for hypothesis testing is the failure of obtaining p-value.

Cochran- Mantel-Haenszel Test

- The idea of CMH test is similar to that of Breslow-Day test: under the null hypothesis,
- n_{k11} is close to its mean $E(n_{k11};1)$ for each k.
As a result, the total $\sum_{k=1}^r n_{k11}$ is also close to its mean, $\sum_{k=1}^r E(n_{k11};1)$

Cochran- Mantel-Haenszel Test

- Cochran- Mantel-Haenszel test statistics takes the form:

$$\chi_{CMH}^2 = \frac{\left[\sum_{k=1}^r n_{k11} - \sum_{k=1}^r E(n_{k11};1) \right]^2}{\sum_{k=1}^r Var(n_{k11};1)}$$

- Under the null hypothesis, Cochran- Mantel-Haenszel test statistics has a chi-squared distribution with degrees of freedom 1.

Cochran- Mantel-Haenszel Test

```
data cmh;
input center smoke cancer count @@;
cards;
1 1 1 126    1 1 0 100    1 0 1 35    1 0 0 61
2 1 1 908    2 1 0 688    2 0 1 497    2 0 0 807
3 1 1 913    3 1 0 747    3 0 1 336    3 0 0 598
4 1 1 235    4 1 0 172    4 0 1 58    4 0 0 121
5 1 1 402    5 1 0 308    5 0 1 121    5 0 0 215
6 1 1 182    6 1 0 156    6 0 1 72    6 0 0 98
7 1 1 60     7 1 0 99     7 0 1 11    7 0 0 43
8 1 1 104    8 1 0 89     8 0 1 21    8 0 0 36
;
run;
proc freq data=cmh order=data;
  weight count;
  table center*smoke*cancer/ cmh ;
run;
```

Cochran- Mantel-Haenszel Test

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	280.1375	<.0001
2	Row Mean Scores Differ	1	280.1375	<.0001
3	General Association	1	280.1375	<.0001

Effect Modifier

- **An effect modifier** is a variable (z) that modifies the association of the exposure variable (x) with the response variable (y). In other words, if z is an effect modifier, then the conditional (adjusted) odds ratio of x and y changes across the level of z.
- For example, y=Alzheimer Disease (AD)
x=gender z=Apoe-4. The association of gender with AD in the apoe-4 group is stronger than that in the non apoe-4 group.

Effect Modifier

- In general, a variable (z) can be classified into four categories according to whether it is a confounder and whether it is an effect modifier:
 1. confounder (yes) and effect modifier (no)
 2. confounder (yes) and effect modifier (yes)
 3. confounder (no) and effect modifier (no)
 4. confounder (no) and effect modifier (yes)

Effect Modifier

- Category 3 & 4 are relevant in clinical trial as z can not be confounder due to randomization even though it can be effect modifier.
- Category 1 & 2 are relevant in observational study as z can be both confounder and effect modifier.

Home Work Assignment

Problem 1 (3.8 on page 68)

Table 3.5 (given on slide 42) refers to the effect of passive smoking on lung cancer. It summarizes results of case-control studies from three countries among nonsmoking women married to smokers. Test the hypothesis that having lung cancer is independent of passive smoking, controlling for country. Report the P-value, and interpret.

(Note: Weak associations in observational studies are suspect. With relatively small changes in the data, perhaps representing effects of misclassification or other bias, the association could disappear. See, for instance, R.L.Tweedie et al., Garbage in, garbage out, *Chance*, 7: no. 2, 20-27(1994))

Problem 2 (3.9 on page 68)

Refer to the previous problem. Assume that the true odds ratio between passive smoking and lung cancer is the same for each study. Estimate its value, and use software to find a 95% confidence interval. Interpret. Analyze whether the odds ratios truly are identical.

Table 3.5

Country	Spouse Smoked	Cases	Controls
Japan	No	21	82
	Yes	73	188
Great Britain	No	5	16
	Yes	19	38
United States	No	71	249
	Yes	137	363