BS852 HW4

Jiayuan Shi

Feb 19, 2016

1

a. First analyze the data ignoring the matching in the study design (Table 1). Report both the odds ratio and the chi-square statistic. What are your conclusions from this analysis?

H0: There is no association between Estrogen Use and the odds of endometrial cancer. OR = 1;
H1: There is an association between Estrogen Use and the odds of endometrial cancer. OR $\neq$ 1.

```
library(epitools)
data <- matrix(c(55, 19, 128, 164), nrow = 2)
rownames(data) <- c("Cases", "Controls")
colnames(data) <- c("Estrogen.Yes", "Estrogen.No")
data
```

```
##          Estrogen.Yes Estrogen.No
## Cases              55         128
## Controls           19         164
```

```
oddsratio.wald(data)
```

```
## $data
##          Estrogen.Yes Estrogen.No Total
## Cases              55         128   183
## Controls           19         164   183
## Total              74         292   366
##
## $measure
##                          NA
## odds ratio with 95% C.I. estimate    lower    upper
##                 Cases    1.000000       NA       NA
##                 Controls 3.708882 2.096434 6.561524
##
## $p.value
##           NA
## two-sided    midp.exact fisher.exact   chi.square
##   Cases              NA           NA           NA
##   Controls 2.334257e-06 3.716595e-06 2.795744e-06
##
## $correction
## [1] FALSE
##
## attr(,"method")
## [1] "Unconditional MLE & normal approximation (Wald) CI"
```

The output gives the odds ratio as 3.71, with the confidence interval of (2.10, 6.56).

```r
chisq.test(data,correct=FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  data
## X-squared = 21.952, df = 1, p-value = 2.796e-06
```

The output gives the chi-squared as 21.952, with the degree of freedom of 1 and p-value = 2.796e-06 < 0.05.

The above p-value of chi-square statistic is < 0.05. Thus, we can reject H0 and this odds ratio is significantly different from 1. The association between Estrogen Use and the odds of endometrial cancer is significant. With 95% confidence, we estimate that the true OR for Cases lies between 2.10 and 6.56. This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of endometrial cancer among those who use menopausal estrogen.

Since the odds ratio is 3.71, we can say that those who use menopausal estrogen have 3.71 times the odds of endometrial cancer, compared to those who do not use menopausal estrogen.

b. Now analyze the data properly, taking the matching into account (Table 2). Report both the odds ratio and the chi-square statistic. What are your conclusions from this analysis?

H0: OR $= 1$;
H1: OR $\neq 1$.

$$OR = \frac{b'}{c'} = \frac{43}{7} = 6.14$$

$$Chi - squared = \frac{(b'-c')^2}{b'+c'} = \frac{(43-7)^2}{43+7} = 25.92(degreeof freedom = 1)$$

The above chi-square statistic is greater than the critical value of 3.84 for 1 degree of freedom for significance level of 0.05, so the p-value is < 0.05. Thus, we can reject H0 and this odds ratio is significantly different from 1. The association between Estrogen Use and the odds of endometrial cancer is statistically significant.

Since the odds ratio is 6.14, we can say that those who use menopausal estrogen have 6.14 times the odds of endometrial cancer, compared to those who do not use menopausal estrogen.

c. Compute a 95% confidence interval for the odds ratio from the matched analysis. Make a statement regarding the association between the use of estrogen and endometrial cancer from this result.

$$95\%CI(OR) = OR^{(1\pm 1.96/\sqrt{\chi^2})} = 6.14^{(1\pm 1.96/\sqrt{25.92})}$$
$$= 6.14^{0.615} to 6.14^{1.384} = 3.05 to 12.33$$

With 95% confidence, we estimate that the true OR for Cases lies between 3.05 and 12.33. This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of endometrial cancer among those who use menopausal estrogen.

d. In which direction is the odds ratio biased in the unmatched analysis? How could this be utilized as an argument for using the unmatched analysis? Why is this argument faulty?

From part a, we get the odds ratio as 3.71 (CI: (2.10, 6.56)); From part b and c, we get the odds ratio as 6.14 (CI: (3.05, 12.33)).The odds ratio biased in the unmatched analysis is in the negative direction, and it is less than the odds ratio in the matched analysis. Using the unmatched analysis, we may get a smaller odds ratio than that with the matched analysis. This argument is faulty, because the matched study uses restriction sampling, and the cases may not typically represent the general population.

2.

a. What portion of the control population are possible matches for person 1 (i.e., same sex and approximate age)? Person 2? Person 3?

For person 1, the possible match is female age from 50-59, and there are 20 females;
For person 2, the possible match is male age from 60-69, and there are 200 males;
For person 3, the possible match is male age from 50-59, and there are 300 males.

b. Assuming the population is randomly mixed, the expected number of persons you have to examine in order to obtain a match for person 1 is 90, and for person 2 is 9.5, and for person 3 is 5.5. Suppose it takes 1/4 hour to examine each person, how long will it take to match person 1? Person 2? Person 3?

For person 1, the expected number of persons to examine to obtain a match is 90, and it takes 1/4 hour to examine a person, so it will take 22.5 (90/4) hours to match.
For person 2, the expected number of persons to examine to obtain a match is 9.5, and it takes 1/4 hour to examine a person, so it will take 2.375 (9.5/4) hours to match.
For person 3, the expected number of persons to examine to obtain a match is 5.5, and it takes 1/4 hour to examine a person, so it will take 1.375 (5.5/4) hours to match.

c. What are your conclusions from this exercise?

Combined part a and part b, the larger portion of possible match for a person, the shorter time it will take to match that person. The smaller portion of possible match for a person, the longer time it takes to match that person.

3

a. Compute a 95% confidence interval for the OR among women under age 70. Interpret this result.

H0: Among women under age 70, there is no association between estrogens and endometrial cancer. OR = 1;
H1: Among women under age 70, there is an association between estrogens and endometrial cancer. OR $\neq$ 1.

Among women under age 70,

$$OR = 7$$

$$Chi - squared = \frac{(b' - c')^2}{b' + c'} = \frac{(7-1)^2}{7+1} = 4.5(degree of freedom = 1)$$

$$95\%CI(OR) = OR^{(1 \pm 1.96/\sqrt{\chi^2})} = 7^{(1 \pm 1.96/\sqrt{4.5})}$$

$$= 7^{0.076} to 7^{1.924} = 1.16 to 42.26$$

With 95% confidence, we estimate that the true OR for Cases lies between 1.16 and 42.26. This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of endometrial cancer among those women under age 70 who use estrogens.

b. Test for interaction of the odds ratios. State your hypothesis and your conclusions in the context of the problem. Show your work.

To test the interaction in matched studies, we build the below 2x2 table *data* and see whether the OR are different.

```
library(epiR)
```

```
## Loading required package: survival
##
## Attaching package: 'survival'
##
## The following object is masked from 'package:epitools':
##
##     ratetable
##
## Package epiR 0.9-69 is loaded
## Type help(epi.about) for summary information
```

```
data <- c(7, 1, 6, 4)
dim(data) <- c(2,2)
dimnames(data)[[1]] <- c("Yes-No", "No-Yes")
dimnames(data)[[2]] <- c("Under 70", "70 and older")
data
```

```
##        Under 70 70 and older
## Yes-No        7            6
## No-Yes        1            4
```

```
epi.2by2(dat=data, method="case.control", homogeneity="breslow.day", conf.level=0.95)
```

```
##              Outcome +    Outcome -      Total        Prevalence *
## Exposed +            7            6         13                53.8
## Exposed -            1            4          5                20.0
## Total                8           10         18                44.4
##                  Odds
## Exposed +        1.17
## Exposed -        0.25
## Total            0.80
##
## Point estimates and 95 % CIs:
## -------------------------------------------------------
## Odds ratio (W)                              4.67 (0.40, 53.95)
## Attrib prevalence (W) *                     33.85 (-10.47, 78.16)
## Attrib prevalence in population (W) *       24.44 (-17.46, 66.35)
## Attrib fraction (est) in exposed  (%)       76.72 (-226.45, 99.62)
## Attrib fraction (est) in population (%)     68.75 (-127.31, 95.70)
```

```
## ------------------------------------------------------
##  X2 test statistic: 1.675 p-value: 0.196
##  W: Wald confidence limits
##  * Cases per 100 population units
```

The OR among women under age 70 is 7.0, and the OR among women 70 and older is 1.5.

H0: OR1 = OR2; H1: OR1 $\neq$ OR2.
From the $OR.homog results, we can get the test statistic chi-square is 1.675, with df=1, and the p-value is 0.196. The p-value is larger than 0.05, so with 95% confidence, we can not reject the null hypothesis. Hence, OR1 and OR2 is similar, and there is no interaction between exogenous estrogens and endometrial cancer matched women using Breslow Day test.

4

a) The MH OR is 5.67, with a chi-square 89.9 and a p-value $<$ 0.0001 that showed evidence of a significant association between smoking and CVD. What is problematic with this analysis?

The analysis ignores the fact that data are matched. There are totally 552 subjects but not 552 pairs, and we cannot use the fomula $\frac{ad}{bc} = \frac{204*184}{72*92} = 5.67$ to calculate MH OR.

b) How many matched pairs are in the analysis?

276 matched pairs are in the analysis. Cases: 276 = 204+72; Controls: 276 = 92+184

c) Complete the table below that reports the number of matched pairs.

	Control	
Case	Former or Current	Never
Former or Current	72	132
Never	20	52

d) Compute the OR to measure the association between smoking and CVD and test whether the association is statistically significant. Use this result to discuss the conclusion of this study. BE COMPLETE

H0: There is no association between smoking and cardio vascular disease. OR = 1;
H1: There is an association between smoking and cardio vascular disease. OR $\neq$ 1.

Incorporating matching,

```
D <- c( rep("E", 204), rep("NE",72))
ND <- c( rep("E",72),   rep("NE",132), rep("E",20), rep("NE",52))
table(D , ND)
```

```
##     ND
## D      E  NE
##   E   72 132
##   NE  20  52
```

```r
mcnemar.test(table(D , ND),correct=F)
```

```
##
##  McNemar's Chi-squared test
##
## data:  table(D, ND)
## McNemar's chi-squared = 82.526, df = 1, p-value < 2.2e-16
```

$$mOR = \frac{b'}{c'} = \frac{132}{20} = 6.6$$

$$Chi - squaredMH = \frac{(b' - c')^2}{b' + c'} = \frac{(132 - 20)^2}{132 + 20} = 82.53(degreeoffreedom = 1)$$

$$95\%CI(OR) = OR^{(1 \pm 1.96/\sqrt{\chi^2})} = 6.6^{(1 \pm 1.96/\sqrt{82.53})}$$

$$= 6.6^{0.784} to 6.6^{1.216} = 4.39 to 9.92$$

The above chi-square statistic 82.53 is greater than the critical value of 3.84 for 1 degree of freedom for significance level of 0.05, so the p-value is < 0.05. Thus, we can reject H0 and this odds ratio is significantly different from 1. The association between smoking and cardio vascular disease is significant. With 95% confidence, we estimate that the true OR for cardio vascular disease lies between 4.39 and 9.92 This confidence interval also excludes 1. There is statistically significant evidence to reject H0 and suggest an increased odds of cardio vascular disease among those former or current smokers.

Since the odds ratio is 6.6, we can say that those former or current smokers have 6.6 times the odds of cardio vascular disease, compared to those who never smoke.