

MA50259: Statistical Design of Investigations

Lab sheet 8: Measures of association in cohort studies

In this practical you will learn about how to adjust for confounding in an observational study. Take the time to run each of the following commands and analyse the displayed results to understand what the code is doing.

Question 1: Hypertension in later life

Some pregnant women are affected by severe gestational hypertension (high blood pressure) known as pre-eclampsia and eclampsia. In a study to investigate the long-term effects of these problems, the proportions of women with hypertension later in life were compared in two groups. One group included women who suffered from gestational hypertension during their first pregnancy. **This is the exposed group.** The other group included women of similar ages to those in the first group, but who did not suffer from these conditions during their first pregnancy. **This is the control group.** Details of the study can be found in [here](#)

The dataset corresponding to this study can be downloaded as an *R* object from the following link

```
mydata<-load(url("http://people.bath.ac.uk/kai21/MA50259/Data/Eclampsia.R"))
```

0. Load the necessary packages in R

```
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

1. Produce the 2 by 2 table corresponding to this study together with the total of women in each exposure group

```
table(Eclampsia$Eclampsia,Eclampsia$Hypertension)
```

```
##
##           No hypertension Yes hypertension
## no eclampsia           201             76
## eclampsia             215            327
```

```
# alternative calculation
cont.table <-
  Eclampsia %>%
  group_by(Eclampsia, Hypertension) %>%
  summarise(n=n()) %>% spread(key = Hypertension, value = n) %>%
  mutate(Total = `No hypertension` + `Yes hypertension`)
print(cont.table)
```

```
## # A tibble: 2 x 4
## # Groups:   Eclampsia [2]
##   Eclampsia   'No hypertension' 'Yes hypertension' Total
##   <fct>             <int>             <int> <int>
## 1 no eclampsia         201                 76   277
## 2 eclampsia           215                327   542
```

2. Produce a table with the total number of women, the estimated probabilities (risks) and the estimated odds of hypertension in later life for each of the two exposure groups

```
table2 <-
  cont.table %>% mutate(Risk=`Yes hypertension`/Total,Odds=`Yes hypertension`/`No hypertension`)
print(table2)
```

```
## # A tibble: 2 x 6
## # Groups:   Eclampsia [2]
##   Eclampsia   'No hypertension' 'Yes hypertension' Total Risk Odds
##   <fct>             <int>             <int> <int> <dbl> <dbl>
## 1 no eclampsia         201                 76   277 0.274 0.378
## 2 eclampsia           215                327   542 0.603 1.52
```

3. Produce point estimates as well as 95% confidence intervals for

- the risk difference

```
p0<-table2[table2$Eclampsia=="no eclampsia","Risk"]
p1<-table2[table2$Eclampsia=="eclampsia","Risk"]
diff<-p1-p0
print(as.numeric(diff))
```

```
## [1] 0.3289528
```

```
summaries.diff<-summary(glm(Hypertension~Eclampsia,data=Eclampsia,
                             family = binomial(link = "identity"))) %>% coefficients()
print(summaries.diff)
```

```
##               Estimate Std. Error   z value    Pr(>|z|)
## (Intercept)    0.2743682 0.02680928 10.234076 1.395201e-24
## Eclampsiaeclampsia 0.3289528 0.03406312  9.657154 4.584219e-22
```

```
print(summaries.diff[2,"Estimate"])
```

```
## [1] 0.3289528
```

```
c(summaries.diff[2,"Estimate"]-1.96*summaries.diff[2,"Std. Error"]  
,summaries.diff[2,"Estimate"]+1.96*summaries.diff[2,"Std. Error"])
```

```
## [1] 0.2621891 0.3957165
```

- the relative risk

```
rr<-p1/p0  
print(rr)
```

```
##      Risk  
## 1 2.198946
```

```
summaries.rr<-summary(glm(Hypertension~Eclampsia,data=Eclampsia,  
                          family = binomial(link = "log"))) %>% coefficients()  
print(summaries.rr)
```

```
##              Estimate Std. Error    z value    Pr(>|z|)  
## (Intercept)   -1.2932842 0.09771277 -13.235569 5.468070e-40  
## Eclampsiaecclampsia 0.7879783 0.10373462  7.596098 3.051933e-14
```

```
print(exp(summaries.rr[2,"Estimate"]))
```

```
## [1] 2.198946
```

```
exp(c(summaries.rr[2,"Estimate"]-1.96*summaries.rr[2,"Std. Error"]  
,summaries.rr[2,"Estimate"]+1.96*summaries.rr[2,"Std. Error"]))
```

```
## [1] 1.794378 2.694730
```

- the odds ratio

```
or<-(p1/(1-p1))/(p0/(1-p0))  
print(or)
```

```
##      Risk  
## 1 4.02246
```

```
summaries.or<-summary(glm(Hypertension~Eclampsia,data=Eclampsia,  
                          family = binomial(link = "logit"))) %>% coefficients()  
print(summaries.or)
```

```
##              Estimate Std. Error    z value    Pr(>|z|)  
## (Intercept)   -0.9725716 0.1346588 -7.222488 5.104487e-13  
## Eclampsiaecclampsia 1.3918937 0.1607553  8.658464 4.781650e-18
```

```
print(exp(summaries.or[2,"Estimate"]))
```

```
## [1] 4.02246
```

```
exp(c(summaries.or[2,"Estimate"]-1.96*summaries.or[2,"Std. Error"]  
,summaries.or[2,"Estimate"]+1.96*summaries.or[2,"Std. Error"]))
```

```
## [1] 2.935311 5.512256
```

4. Perform the test of hypothesis

$$H_0 : p_1 = p_0 \quad \text{vs} \quad H_1 : p_1 \neq p_0$$

using the χ^2 test and also the normal approximation to the sampling distribution of the estimated

- risk difference
- relative risk
- odds ratio

```
chisq.test(x=Eclampsia$Eclampsia,y=Eclampsia$Hypertension,correct=F)
```

```
##  
## Pearson's Chi-squared test  
##  
## data: Eclampsia$Eclampsia and Eclampsia$Hypertension  
## X-squared = 79.366, df = 1, p-value < 2.2e-16
```

```
summaries.diff[2,"Pr(>|z|)"]
```

```
## [1] 4.584219e-22
```

```
summaries.rr[2,"Pr(>|z|)"]
```

```
## [1] 3.051933e-14
```

```
summaries.or[2,"Pr(>|z|)"]
```

```
## [1] 4.78165e-18
```

Question 2: Treatment for kidney stones

Kidney stones can cause intense pain, obstruct the urinary tract or damage the kidneys. In the 1980s several new techniques were introduced to remove kidney stones, in preference to the then standard form of treatment which involved open surgery. These new techniques included keyhole surgery and shockwave therapy.

A study was undertaken to compare the different treatment methods. This exercise focuses on open surgery and keyhole surgery (the technical term is *percutaneous nephrolithotomy*).

The treatment was defined as successful if the stones were eliminated or reduced to less than 2mm after three months. The study was a cohort study, including 350 patients treated with open surgery and 350 with keyhole surgery. In this study the exposure is the new treatment, namely keyhole surgery, and the “*disease*” outcomes are success or failure of the treatment. Information on the size of the stones treated, small (less than 2 cm diameter) or large (more than 2cm diameter), was also collected.

Details of the study can be found [here](http://people.bath.ac.uk/kai21/MA50259/Data/Kidney.R)

The dataset corresponding to this study can be loaded while in *R* by typing the following command

```
load(url("http://people.bath.ac.uk/kai21/MA50259/Data/Kidney.R"))
```

1. Produce the 2 by 2 table corresponding to the association between the exposure and the disease (response) with the corresponding totals for each treatment group

Solution

```
cont.table <-
  Kidney %>%
  group_by(Treatment, Success) %>%
  summarise(n=n()) %>% spread(key = Success, value = n) %>% mutate(Total =Failure+Success)
print(cont.table)
```

```
## # A tibble: 2 x 4
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total
##   <fct>          <int>   <int> <int>
## 1 Keyhole surgery     61     289   350
## 2 Open surgery       77     273   350
```

2. Obtain the odds (of success) ratio for both treatment groups and then obtain the corresponding odds ratio and its corresponding 95% confidence interval.

Solution

```
table2 <-
  cont.table %>% mutate(Odds=Success/Failure)
print(table2)
```

```
## # A tibble: 2 x 5
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total Odds
##   <fct>          <int>   <int> <int> <dbl>
## 1 Keyhole surgery     61     289   350  4.74
## 2 Open surgery       77     273   350  3.55
```

```
OR<-table2[1,"Odds"]/table2[2,"Odds"]
print(OR)
```

```
##      Odds
## 1 1.336276
```

```
Kidney$Treatment<-relevel(Kidney$Treatment,ref='Open surgery')
summaries<-summary(glm(Success~Treatment,data=Kidney,family=binomial))>% coefficients()

print(summaries)
```

```
##              Estimate Std. Error  z value    Pr(>|z|)
## (Intercept)    1.2656664   0.1290349  9.808716 1.032742e-22
## TreatmentKeyhole surgery 0.2898865   0.1910593  1.517259 1.292012e-01
```

```
print(exp(summaries[2,"Estimate"]))
```

```
## [1] 1.336276
```

```
exp(c(summaries[2,"Estimate"]-1.96*summaries[2,"Std. Error"]
,summaries[2,"Estimate"]+1.96*summaries[2,"Std. Error"]))
```

```
## [1] 0.9188892 1.9432516
```

- For each of the two strata defined by the size of the stones treated, produce the 2 by 2 table corresponding to the association between the exposure and the disease with the corresponding totals for each treatment group. Also obtain the corresponding the odds ratio and its corresponding 95% confidence interval for each stratum.

```
cont.table.small <-
  Kidney %>% filter(Size=="Small") %>%
  group_by(Treatment, Success) %>%
  summarise(n=n()) %>% spread(key = Success, value = n) %>% mutate(Total =Failure+Success)
```

```
## 'summarise()' has grouped output by 'Treatment'. You can override using the
## '.groups' argument.
```

```
print(cont.table.small)
```

```
## # A tibble: 2 x 4
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total
##   <fct>          <int>   <int> <int>
## 1 Open surgery      6      81    87
## 2 Keyhole surgery  36     234   270
```

```
table.small<-cont.table.small %>% mutate(Odds=Success/Failure)
print(table.small)
```

```
## # A tibble: 2 x 5
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total Odds
##   <fct>          <int>   <int> <int> <dbl>
## 1 Open surgery      6      81    87  13.5
## 2 Keyhole surgery  36     234   270   6.5
```

```
OR.small<-table.small[2,"Odds"]/table.small[1,"Odds"]
print(OR.small)
```

```
##           Odds
## 1 0.4814815
```

```
summaries.small<-summary(glm(Success~Treatment,data=Kidney,
                             family=binomial,subset=Size=="Small"))>% coefficients()

print(exp(summaries.small[2,"Estimate"]))
```

```
## [1] 0.4814815
```

```
exp(c(summaries.small[2,"Estimate"]-1.96*summaries.small[2,"Std. Error"]
,summaries.small[2,"Estimate"]+1.96*summaries.small[2,"Std. Error"])))
```

```
## [1] 0.1956667 1.1847922
```

```
cont.table.large <-
  Kidney %>% filter(Size=="Large") %>%
  group_by(Treatment, Success) %>%
  summarise(n=n()) %>% spread(key = Success, value = n) %>% mutate(Total =Failure+Success)
```

```
## 'summarise()' has grouped output by 'Treatment'. You can override using the
## '.groups' argument.
```

```
print(cont.table.large)
```

```
## # A tibble: 2 x 4
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total
##   <fct>          <int>   <int> <int>
## 1 Open surgery      71     192   263
## 2 Keyhole surgery   25      55    80
```

```
table.large<-cont.table.large %>% mutate(Odds=Success/Failure)
```

```
print(table.large)
```

```
## # A tibble: 2 x 5
## # Groups:   Treatment [2]
##   Treatment      Failure Success Total Odds
##   <fct>          <int>   <int> <int> <dbl>
## 1 Open surgery      71     192   263  2.70
## 2 Keyhole surgery   25      55    80  2.2
```

```
OR.large<-table.large[2,"Odds"]/table.large[1,"Odds"]
print(OR.large)
```

```
##          Odds
## 1 0.8135417
```

```
summaries.large<-summary(glm(Success~Treatment,data=Kidney,
                             family=binomial,subset=Size=="Large"))%>% coefficients()

print(exp(summaries.large[2,"Estimate"]))
```

```
## [1] 0.8135417
```

```
exp(c(summaries.large[2,"Estimate"]-1.96*summaries.large[2,"Std. Error"]
,summaries.large[2,"Estimate"]+1.96*summaries.large[2,"Std. Error"]))
```

```
## [1] 0.4714654 1.4038147
```

4. What is the evidence of confounding by the size of the kidney stones?

Solution

```
summaries<-summary(glm(Success~Treatment+Size,data=Kidney,family=binomial))%>% coefficients()

print(exp(summaries[2,"Estimate"]))
```

```
## [1] 0.6996125
```

```
exp(c(summaries[2,"Estimate"]-1.96*summaries[2,"Std. Error"]
,summaries[2,"Estimate"]+1.96*summaries[2,"Std. Error"]))
```

```
## [1] 0.4465411 1.0961089
```

The regression summary odds ratio is about 0.7 while the crude OR is about 1.33 then we have evidence of confounding.