

# Statistics 305/605: Introduction to Biostatistical Methods for Health Sciences

R Demo for Chapter 16: Multiple  $2 \times 2$  Contingency Tables

Jinko Graham

2018-10-28

## Example Data: Smoking, Aortic Stenosis and Sex

```
uu <- url("http://people.stat.sfu.ca/~jgraham/Teaching/S305_17/Data/sten.txt")
stenosis <- read.table(uu,header=TRUE)
head(stenosis)
```

```
##      smoke AS sex
## 1         1  1  1
## 2         1  1  1
## 3         1  1  1
## 4         1  1  1
## 5         1  1  1
## 6         1  1  1
```

- ▶ AS is 1 for patients with aortic stenosis (AS) and 0 otherwise.
- ▶ smoke is 1 for smokers and 0 for non-smokers
- ▶ sex is 1 for males and 0 for females.

# Tabulating data in multiway tables

- Can use the `xtabs()` function to get tables of AS by smoke that are stratified by sex:

```
xtabs(~AS + smoke + sex, data=stenosis)
```

```
## , , sex = 0
```

```
##
```

```
##      smoke
```

```
## AS    0  1
```

```
##      0 47 19
```

```
##      1 29 14
```

```
##
```

```
## , , sex = 1
```

```
##
```

```
##      smoke
```

```
## AS    0  1
```

```
##      0 20 24
```

```
##      1 25 37
```

- ▶ Tables of AS by smoke without regard to sex
  - ▶ This is called the “pooled table”, as opposed to the stratified tables from the previous slide.

```
xtabs(~AS + smoke, data=stenosis)
```

```
##      smoke
## AS    0  1
##    0 67 43
##    1 54 51
```

# Software Notes

- ▶ `xtabs()` uses R formulas; e.g., `~ AS + smoke + sex`.
- ▶ The order of the variables in the formula specifies the order of the margins.
  - ▶ If you switched to, say, `~ AS + sex + smoke`, you would have tables of disease by sex that are stratified by smoking status.

# Test of Homogeneity in the AS Example

- ▶ Are the ORs for AS the same across the different sexes? Does sex modify the effect of smoking on AS?
- ▶ R has no built-in function or add-on packages (that I know of) for this test. You can use one that I wrote. Below are the steps.
  1. For each stratum  $i$  of the extraneous variable, get the stratum-specific log-ORs and weights,  $y_i = \log_e(\widehat{OR}_i)$ 's and  $w_i = 1/se^2(\log_e \widehat{OR}_i)$  using `stratumStats()`
  2. Pass these stratum-specific statistics to `homogTest()`, which combines them into the homogeneity test statistic and returns a  $p$ -value for the test of homogeneity of ORs across the strata.

## stratumStats()

- Get the stratum-specific log-OR,  $y_i = \log_e(\widehat{OR}_i)$ , and weight  $w_i = 1/se^2(\log_e \widehat{OR}_i)$ ,

```
stratumStats <- function(a,b,c,d) {  
  y<-log(a*d/(b*c))  
  w <- 1/(1/a+1/b+1/c+1/d)  
  return(list(y=y,w=w))  
}
```

- Apply stratumStats() to the first stratum defined by the extraneous variable sex (for females).

```
s1 <- stratumStats(47,19,29,14)  
s1
```

```
## $y  
## [1] 0.1774701  
##  
## $w  
## [1] 5.561132
```

- ▶ Then apply `stratumStats()` to the 2nd stratum defined by the sex (for males).

```
s2 <- stratumStats(20,24,25,37)
s2
```

```
## $y
## [1] 0.2097205
##
## $w
## [1] 6.301448
```

- ▶ The R objects `s1` and `s2` contain the  $(y_i, w_i)$  pairs for the 1st (females) and 2nd (males) stratum of sex, respectively.



## homogTest()

- ▶ `homogTest()` takes the stratum-specific log-ORs and weights in `s1` and `s2` as arguments and returns the  $X^2$  statistic for homogeneity, along with a  $p$ -value.

```
homogStat <- function(s1, s2) {  
  Y <- (s1$w*s1$y + s2$w*s2$y)/(s1$w+s2$w)  
  X2 <- s1$w*(s1$y-Y)^2 + s2$w*(s2$y-Y)^2  
  p <- pchisq(X2,df=1,lower.tail=FALSE)  
  return(list(X2=X2,pval=p))  
}
```

- ▶ Apply `homogStat()`:

```
homogStat(s1,s2)
```

```
## $X2  
## [1] 0.003072521  
##  
## $pval  
## [1] 0.9557956
```

- ▶ We retain the hypothesis of homogeneous ORs. Insufficient evidence to claim that sex modifies the effect of smoking on AS.

# Confounding

## Simpson's Paradox and the AS Example

- ▶ As sex doesn't appear to be an effect modifier, we can consider whether it is a confounder.
- ▶ Do this informally, by checking whether the ORs when we account for sex are different from the OR when we ignore sex.

```
OR1<-exp(s1$y) #OR in females  
OR1
```

```
## [1] 1.194192
```

```
OR2<-exp(s2$y) #OR in males  
OR2
```

```
## [1] 1.233333
```

- ▶ The estimated ORs for smoking and AS, stratified by sex are about the same: 1.19 for females and 1.23 for males.
  - ▶ This suggests that, for both females and males, smoking *slightly* increases the odds of AS.

- ▶ However, the estimated OR from the pooled table (when we ignore sex) is 1.47.
  - ▶ This suggests a *stronger* effect of smoking than we obtain from the stratified tables!

```
xtabs(~AS + smoke, data=stenosis)
```

```
##      smoke
## AS    0   1
##      0 67 43
##      1 54 51
```

```
a<-67; b<-43; c<-54; d<-51
ORpooled<-a*d/(b*c)
ORpooled
```

```
## [1] 1.471576
```

- ▶ Simpson's paradox: When the effect of an exposure on the odds of the disease outcome differs in the stratified and pooled analyses.

# The MH Common OR Estimate and Test

- ▶ The effect of smoking on AS appears to differ between the analyses that stratify on versus ignore sex.
- ▶ We therefore adjust for sex as a potential confounding variable by getting the MH estimate of the common (adjusted) OR.
- ▶ Once we have the MH estimate of the common OR across the strata of the extraneous variable, we may test  $H_0$  that this common OR is one.
- ▶ The text, section 16.2.3, goes into the details of the derivation, if you're interested.
- ▶ Instead, focus on using R to apply this test and interpreting the results.
  - ▶ R has a built-in function `mantelhaen.test()` that we can apply.

# MH Test applied to the AS Example

- ▶ Apply `mantelhaen.test()` to the stenosis data.
  - ▶ As the text doesn't use a continuity correction, to be consistent, we turn off the default setting in the function.

```
with(stenosis,mantelhaen.test(x=smoke,y=AS,z=sex,correct=FALSE,  
                             conf.level=0.95))
```

```
##  
## Mantel-Haenszel chi-squared test without continuity correction  
##  
## data:  smoke and AS and sex  
## Mantel-Haenszel X-squared = 0.44568, df = 1, p-value = 0.5044  
## alternative hypothesis: true common odds ratio is not equal to 1  
## 95 percent confidence interval:  
##  0.6877312 2.1460496  
## sample estimates:  
## common odds ratio  
##           1.214868
```

- ▶ Referring to the p-value reported in the output, we see that, at the 5% level, there is no statistical evidence that smoking affects the odds of AS, once we adjust for sex.