

Stats 111 Homework 2

Owen Lin

```
## Warning: package 'pROC' was built under R version 4.2.2
```

```
## Type 'citation("pROC")' for a citation.
```

```
##
## Attaching package: 'pROC'
```

```
## The following objects are masked from 'package:stats':
##
##      cov, smooth, var
```

1. a. H_0 : Cholesterol level and CHD are independent
 H_a : Cholesterol level and CHD are dependent
- b. test statistic: 35.028 ~ chisquared(3)
p-value: $1.202 * 10^{-7}$

```
heart = matrix(c(307, 246, 439, 245, 12, 8, 31, 41), 4, 2)
rownames(heart) = c("low", "above normal", "high", "very high")
colnames(heart) = c("CHD no", "CHD yes")

# chi square test of independence
chisq.test(heart , correct=FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data: heart
## X-squared = 35.028, df = 3, p-value = 1.202e-07
```

- c. conclusion: Cholesterol level and CHD are dependent at 5% significant level
- d. It is natural to think that higher cholesterol tends to lead to heart disease. So “low”, “above normal”, “high”, “very high” comes in a natural order.
- e. H_0 : Probability of CHD didn't increase with Cholesterol level.
 H_a : Probability of CHD increases with Cholesterol level.
- f. test statistic: 26.167 ~ chisquared(1)
p-value: $3.131 * 10^{-7}$
conclusion: We reject the null and conclude that there is a trend between CHD and Cholesterol level.

```
# addmargins(heart)
prop.trend.test(heart[,2], rowSums(heart))
```

```
##
## Chi-squared Test for Trend in Proportions
##
## data: heart[, 2] out of rowSums(heart) ,
## using scores: 1 2 3 4
## X-squared = 26.167, df = 1, p-value = 3.131e-07
```

2. a. I use chi-squared of independence and got a p-value $< 2.2 * 10^{-16}$. So there is evidence that socioeconomic level is correlated with delinquent status.

```
# This is the 3 way scouts table
scout = array( c(169,43,42,11,59,196,10,10),dim = c(2, 2, 2),dimnames = list( Scout = c("No","Yes"),Verdict = c("Delinquent No", "Delinquent Yes"),Socioeconomic = c("Low","High")))

# This is the two way table of Socioeconomic status and Scout status
socio_scout = array(c(211,69,54,206), dim=c(2,2),dimnames=list(Socioeconomic=c("Low","High"), Scout=c("No","Yes")))

chisq.test(socio_scout , correct=FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data: socio_scout
## X-squared = 160.75, df = 1, p-value < 2.2e-16
```

- b. I use odds ratio estimate and got a odds ratio of 0.314. So there is evidence that socioeconomic level is correlated with scout status.

```
# This is the two way table of Socioeconomic stats and Delinquency status
socio_deliq = array(c(212,255,53,20), dim=c(2,2),dimnames=list(Socioeconomic=c("Low","High"), Deliquent=c("No","Yes")))

prop.comp(socio_deliq)
```

```
## Warning in prop.comp(socio_deliq): Estimated probabilities represent Pr[ Deliquent | Socioeconomic ]. For estimates of
## Pr[ Socioeconomic | Deliquent ], change the value of 'transpose'.
```

```
## $riskdiff
##           Deliquent
## Socioeconomic No      p0 Yes      p1 riskdiff      lower      upper
##           Low  212 0.8000000  53 0.20000000 0.0000000      NA      NA
##           High 255 0.9272727  20 0.07272727 -0.1272727 -0.1880864 -0.06645901
##           Deliquent
## Socioeconomic      p.value
##           Low      NA
##           High 1.531149e-05
##
## $riskratio
##           Deliquent
## Socioeconomic No      p0 Yes      p1 riskratio      lower      upper
##           Low  212 0.8000000  53 0.20000000 1.0000000      NA      NA
##           High 255 0.9272727  20 0.07272727 0.3636364 0.2236909 0.5911345
##           Deliquent
## Socioeconomic      p.value
##           Low      NA
##           High 1.531149e-05
##
## $oddsratio
##           Deliquent
## Socioeconomic No      p0 Yes      p1 oddsratio      lower      upper
##           Low  212 0.8000000  53 0.20000000 1.0000000      NA      NA
##           High 255 0.9272727  20 0.07272727 0.3137255 0.1817923 0.5414074
##           Deliquent
## Socioeconomic      p.value
##           Low      NA
##           High 1.531149e-05
```

- c. Socioeconomic status looks like a confounder in the association between Scout status and Delinquency status. Because it is related to both Scout and Delinquency Status.
- d. $OR_{MH} = 0.657$, it's rather away from 1, so it seems that delinquent and scout status are dependent when controlling for socioeconomic level.

```
OR_MH = (169*11/(169+42+43+11)+59*10/(59+10+196+10))/(43*42/(169+42+43+11)+196*10/(59+10+196+10))
```

- e. $H_0: OR_{MH} = 1$ $H_a: OR_{MH} \neq 1$ p-value: 0.1843 conclusion: we fail to reject the null => ie. delinquent and scout status are uncorrelated when controlling for socioeconomic levels.

```
mantelhaen.test(scout)
```

```
##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data:  scout
## Mantel-Haenszel X-squared = 1.7622, df = 1, p-value = 0.1843
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.3635458 1.1874379
## sample estimates:
## common odds ratio
##          0.6570297
```

f. The effect of being delinquent on scout status differs across different socioeconomic levels. So socioeconomic level is a effect modifier.

3. a. If the person doesn't drink and smoke, then the expected BP is beta 0. It is a meaningful interpretation because it is a plausible condition.

b. If the person Smoke, then drink brings additional beta 3 effect on BP.

c. The effect of drinking on BP for smokers is $\beta_2 + \beta_3$ more than nonsmokers.

d. $H_0: \beta_2 = \beta_3 = 0$

$H_a: H_0$ is not true

Reduced Model: $y = \beta_0 + \beta_1 \text{Drink}$

Full Model: $y = \beta_0 + \beta_1 \text{Drink} + \beta_2 \text{Smoke} + \beta_3 \text{Drink} \times \text{Smoke}$

4. a. BMI reasonably correlates with blood pressure and the prevalence of CHD.

b. Both p-values are way below 0.05 significant level. That means BMI is a confounder.

H_0 : SBP and BMI are independent.

H_a : SBP and BMI are dependent.

H_0 : CHD and BMI are independent.

H_a : CHD and BMI are dependent.

```
# Framingham Data
framingham = read.table("C:\\Users\\Owen\\Downloads\\Framingham copy.txt")
##### Recode sex to something obvious (sex=1 -> male)
framingham$sex = framingham$sex - 1
names( framingham )[1] = "female"

##### Create SBP and BMI groups
framingham$sbphi = cut( framingham$sbp, breaks=c(min(framingham$sbp),146, max(framingham$sbp)),
include.lowest=TRUE )
framingham$bmigrp = cut( framingham$bmi, breaks=c(min(framingham$bmi, na.rm=TRUE),20, 25, 30, ma
x(framingham$bmi, na.rm=TRUE)), include.lowest=TRUE, right=FALSE )

bmisbp.table = xtabs( ~ bmigrp + sbphi, data=framingham )
epitab( bmisbp.table, pvalue="chi2" )
n.hisbp = bmisbp.table[,2]
n.strata = rowSums(bmisbp.table)
chisq.test(bmisbp.table)
prop.trend.test( n.hisbp, n.strata )

##### Compute test of independence and test for trend for BMI and CHD
bmichd.table = xtabs( ~ bmigrp + chdfate, data=framingham )
epitab( bmichd.table, pvalue="chi2" )
n.chd = bmichd.table[,2]
n.strata = rowSums(bmichd.table)
chisq.test(bmichd.table)
prop.trend.test(n.chd, n.strata)
```

```
## $tab
##           sbphi
## bmigrp    [80,146]      p0 (146,270]      p1 oddsratio      lower
## [16.2,20)      231 0.06292563      24 0.0235525  1.000000      NA
## [20,25)      1685 0.45900300      283 0.2777233  1.616543 1.042364
## [25,30)      1399 0.38109507      467 0.4582924  3.212920 2.083356
## [30,57.6]      356 0.09697630      245 0.2404318  6.623947 4.220519
##           sbphi
## bmigrp      upper      p.value
## [16.2,20)      NA      NA
## [20,25)      2.507005 3.048828e-02
## [25,30)      4.954915 2.938902e-08
## [30,57.6] 10.396037 1.606474e-19
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "chi2"
##
##
## Pearson's Chi-squared test
##
## data:  bmisbp.table
## X-squared = 225.24, df = 3, p-value < 2.2e-16
##
##
## Chi-squared Test for Trend in Proportions
##
## data:  n.hisbp out of n.strata ,
## using scores: 1 2 3 4
## X-squared = 214.6, df = 1, p-value < 2.2e-16
##
## $tab
##           chdfate
## bmigrp      0      p0  1      p1 oddsratio      lower      upper
## [16.2,20)  220 0.06836544  35 0.02377717  1.000000      NA      NA
## [20,25)  1465 0.45525171 503 0.34171196  2.158167 1.489560 3.126885
## [25,30)  1180 0.36668738 686 0.46603261  3.654237 2.526936 5.284444
## [30,57.6]  353 0.10969546 248 0.16847826  4.416026 2.984212 6.534820
##           chdfate
## bmigrp      p.value
## [16.2,20)      NA
## [20,25)  3.307489e-05
## [25,30)  3.227521e-13
## [30,57.6] 4.767196e-15
##
## $measure
## [1] "wald"
```

```
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "chi2"
##
##
## Pearson's Chi-squared test
##
## data:  bmichd.table
## X-squared = 120.25, df = 3, p-value < 2.2e-16
##
##
## Chi-squared Test for Trend in Proportions
##
## data:  n.chd out of n.strata ,
## using scores: 1 2 3 4
## X-squared = 114.14, df = 1, p-value < 2.2e-16
```

c. OR for CHD against BMI in the level of [20,25) is 2.16 with a 95% CI of (1.49, 3.13)

OR for BMI in the level of [25,30) is 3.65 with a 95% CI of (2.53, 5.28).

OR for BMI in the level of [30,57.6] is 4.42 with a 95% CI of (2.98, 6.53).

d. The MH common odds ratio is 1.66 with a 95% CI of (1.43,1.92). Because 1 is not in the CI, so CHD and sbp are related while adjusting for the level of BMI.

```
##### Compute relevant counts
n.sbplo = xtabs( ~ sbphi + bmigrp, data=framingham )[1,]
n.sbphi = xtabs( ~ sbphi + bmigrp, data=framingham )[2,]
lo.chd = xtabs( ~ chdfate + sbphi + bmigrp, data=framingham )[2,1,]
hi.chd = xtabs( ~ chdfate + sbphi + bmigrp, data=framingham )[2,2,]

# Just to show the 3 way table that is being used above for the MH and BD tests
sbphi.chd.table = table(BP=framingham$sbphi, CHD=framingham$chdfate, BMI=framingham$bmigrp)

##### Compute M-H estimate of adjusted OR and test for heterogeneity
mh.rslt = meta.MH(n.sbphi, n.sbplo, hi.chd, lo.chd, names=levels(framingham$bmigrp))
summary( mh.rslt )
```

```
## Fixed effects ( Mantel-Haenszel ) meta-analysis
## Call: meta.MH(ntrt = n.sbphi, nctrl = n.sbplo, ptrt = hi.chd, pctrl = lo.chd,
##      names = levels(framingham$bmigrp))
## -----
##              OR (lower  95% upper)
## [16.2,20) 3.78    1.48    9.66
## [20,25)   1.84    1.41    2.40
## [25,30)   1.55    1.25    1.92
## [30,57.6] 1.57    1.13    2.18
## -----
## Mantel-Haenszel OR =1.66 95% CI ( 1.43,1.92 )
## Test for heterogeneity: X^2( 3 ) = 4.03 ( p-value 0.2579 )
```

```
# MH test using 3 way table
mantelhaen.test(sbphi.chd.table)
```

```
##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data:  sbphi.chd.table
## Mantel-Haenszel X-squared = 45.864, df = 1, p-value = 1.268e-11
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.433431 1.924566
## sample estimates:
## common odds ratio
##          1.660943
```

e. $H_0: OR_{MH} = 1$

$H_a: OR_{MH} \neq 1$

p-value: $1.268 * 10^{-11}$

conclusion: We reject the null and conclude that CHD and sbp are dependent while adjusting for BMI.