Stats 111 HW2

Viraj Vijaywargiya

2023-02-02

```
library(epitools)
library(rmeta)
library(pROC)
```

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

##
## Attaching package: 'pROC'

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

```
library(nnet)
```

1. 1a) The Null hypothesis will be H0 : pij = pi pj for all i=1,2,3,4 and j=1,2. That is CHD and Cholesterol level group are independent.

   Alternative Hypothesis HA: at least one combination of i and j has pij != pi pj. That is CHD and Cholesterol level group are not independent.

   1b)

   ```
   chol = matrix(c(307,246,439,245,12,8,31,41),4,2)
   chisq.test(chol)
   ```

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

   Test statistic: 35.028, P-value: 1.202e-07.

   1c) The p-value < 0.05, therefore, we would reject the null and conclude evidence for the alternative, that CHD and Cholesterol level group are dependent. Hence, we have evidence that Cholesterol level is associated with CHD status.

   1d) The levels of cholesterol can be considered as an ordinal variable because it can have a defined order of ranking. The categories have a meaningful order and can be ranked as "Normal", "Above Normal", "High", "Very High". The fact that high cholesterol is generally accepted as leading to heart disease supports the meaningful order of the ranges.

1e) The null hypothesis is H0 : pi1 = pi2 = pi3 = pi4 (CHD and Cholesterol level group are independent) The alternative hypothesis is Ha : pi1 > pi2 > pi3 > pi4 or Ha : pi1 < pi2 < pi3 < pi4 (CHD and Cholesterol level group are dependent, and there is a trend in the effect of Cholesterol level on CHD fate).

1f)

```
cholesterol = array(c(307,246,439,245,12,8,31,41), dim=c(4,2), dimnames=list( Cholesterol=c("Normal
n.chol = cholesterol[,2]
n.strata = rowSums(cholesterol)
prop.trend.test(n.chol, n.strata )
```

```
##
##  Chi-squared Test for Trend in Proportions
##
## data:  n.chol out of n.strata ,
##  using scores: 1 2 3 4
## X-squared = 26.167, df = 1, p-value = 3.131e-07
```

The p-value < 0.05, therefore, we would reject the null and conclude evidence for the alternative, that CHD and Cholesterol level group are dependent. Hence, we have evidence that the probability of CHD increases or decreases as we move along the Cholesterol level group.

2.
```
scout = array( c(169,43,42,11,59,196,10,10),dim = c(2, 2, 2),dimnames = list( Scout = c("No","Yes")
```

2a) Null Hypothesis H0: Scout status and Socioeconomic level are independent.

Alternate Hypothesis HA: Scout status and Socioeconomic level are not independent.

Using the chi-squared test of independence,

```
socio = array(c(211,69,54,206), dim=c(2,2),dimnames=list(Socioeconomic=c("Low","High"), Scout=c("No
chisq.test(socio)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  socio
## X-squared = 158.57, df = 1, p-value < 2.2e-16
```

P-value $<$ 2.2e-16. Therefore, we reject the null hypothesis and conclude the alternate, that is Scout status and Socioeconomic level are not independent. Thus, Socioeconomic level is associated with Scout status.

2b) Null Hypothesis H0: Delinquency status and Socioeconomic level are independent.

Alternate Hypothesis HA: Delinquency status and Socioeconomic level are not independent.

Using the chi-squared test of independence,

```
socio.deliq = array(c(212,255,53,20), dim=c(2,2),dimnames=list(Socioeconomic=c("Low","High"), Delic
chisq.test(socio.deliq)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  socio.deliq
## X-squared = 17.626, df = 1, p-value = 2.689e-05
```

P-value = 2.689e-05. Therefore, we reject the null hypothesis and conclude the alternate, that is Delinquency status and Socioeconomic level are not independent. Thus, Socioeconomic level is associated with Delinquency status.

2c) Socioeconomic status can be a potential confounder in the association between Scout status and Delinquency status because it is associated with both Scout status and Delinquency status.

2d) Mantel-Haenszel common odds ratio: 0.6570297. Therefore, the common odds ratio is not equal to 1.

2e) The null hypothesis is H0 : OR(MH) = 1. The alternative hypothesis Ha : OR(MH) != 1.

```
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
```

P-value = 0.1843. Since the p-value > 0.05, we fail to reject the null hypothesis.

2f) Socioeconomic status is an effect modifier for the association between scout status and delinquency status because it changes the strength and direction of the relationship. For children from low socioeconomic backgrounds, being a scout is associated with a slightly higher odds of being delinquent (odds ratio of 1.03), while for children from high socioeconomic backgrounds, being a scout is associated with a significantly lower odds of being delinquent (odds ratio of 0.30). This indicates that the association between scout status and delinquency status is influenced by the child's socioeconomic background.

3. 3a) The estimate B0 is the intercept of the regression model, representing the expected mean systolic blood pressure (BP) when both alcohol drinking status and tobacco smoking status are equal to 0. In other words, it represents the baseline systolic blood pressure in the population for individuals who don't drink alcohol and don't smoke tobacco. This is a meaningful interpretation, as it provides a reference point for comparing the effect of alcohol drinking and tobacco smoking on systolic blood pressure.

   3b) The B3 term is the coefficient of the interaction between alcohol drinking status (X1/Drink) and tobacco smoking status (X2/Smoke). It represents the additional effect of alcohol drinking on systolic blood pressure that is specific to individuals who smoke tobacco. The B3 term provides information on how the association between alcohol drinking and systolic blood pressure is modified by tobacco smoking status. For example, if B3 is positive and significantly different from zero, it suggests that the effect of alcohol drinking on systolic blood pressure is stronger for individuals who smoke tobacco than for those who don't.

   3c) For non-smokers (X2/Smoke = 0), the expected systolic blood pressure for a person who drinks (X1/Drink = 1) can be calculated as BP = B0^ + B1^ *1 + B2^ *0 + B3^ *1 *0 + E = B0^ + B1^ + E. This equation suggests that the effect of drinking on systolic blood pressure for non-smokers is given by B1^.

   For smokers (X2/Smoke = 1), the expected systolic blood pressure for a person who drinks (X1/Drink = 1) can be calculated as BP = B0^ + B1^ *1 + B2^ *1 + B3^ *1 *1 + E = B0^ + B1^ + B2^ + B3^ + E. This equation suggests that the effect of drinking on systolic blood pressure for smokers is given by B1^ + B2^ + B3^.

In conclusion, the effect of drinking on systolic blood pressure is different for smokers and non-smokers and is given by B1^ + B2^ + B3^ for smokers and B1^ for non-smokers.

3d) Null hypothesis H0: B2 = B3 = 0. The effect of tobacco smoking (X2/Smoke) on systolic blood pressure (BP) is equal to zero, both as a main effect and as an interaction with alcohol drinking status (X1/Drink).

Alternate hypothesis HA: B2 != 0 or B3 != 0. The effect of tobacco smoking (X2/Smoke) on systolic blood pressure (BP) is not equal to zero, either as a main effect or as an interaction with alcohol drinking status (X1/Drink).

Model 1: BP(i) = B0 + B1Drink(i) + B2Smoke(i) + E(i)

Model 2: BP(i) = B0 + B1Drink(i) + B2Smoke(i) + B3Drink(i)Smoke(i) + E(i)

4.
```
framingham = read.table("/Users/virajvijaywargiya/Downloads/framingham.txt")

framingham$sex = framingham$sex - 1
names( framingham )[1] = "female"
framingham[1:10,]
```

```
##    female sbp dbp scl chdfate followup age  bmi month   id
## 1       0 120  80 267       1       18  55 25.0     8 2642
## 2       0 130  78 192       1       35  53 28.4    12 4627
## 3       0 144  90 207       1      109  61 25.1     8 2568
## 4       0  92  66 231       1      147  48 26.2    11 4192
## 5       0 162  98 271       1      169  39 28.4    11 3977
## 6       1 212 118 182       1      199  61 33.3     2  659
## 7       0 140  85 276       1      201  44 25.3     6 2290
## 8       0 174 102 259       1      209  39 27.9    11 4267
## 9       0 142  94 242       1      265  47 26.6     5 2035
## 10      0 115  70 242       1      278  60 30.8    10 3587
```

```
framingham$sbphi = cut( framingham$sbp, breaks=c(min(framingham$sbp),146, max(framingham$sbp)), inc
```

4a) BMI is a possible cofounder in the relationship between systolic blood pressure and the prevalence of CHD because it could independently affect both systolic blood pressure and CHD. A higher BMI is associated with both higher blood pressure and increased risk of CHD, so if not controlled for, it could artificially inflate the observed association between systolic blood pressure and CHD.

4b)

```
framingham$bmigrp = cut( framingham$bmi, breaks=c(min(framingham$bmi, na.rm=TRUE),20, 25, 30, max(f
```

Test of independence and test for trend for association between BMI and SBP,

```
bmisbp.table = xtabs( ~ bmigrp + sbphi, data=framingham )
epitab( bmisbp.table, pvalue="chi2" )
```

```
## $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"
```

```
n.hisbp = bmisbp.table[,2]
n.strata = rowSums(bmisbp.table)
chisq.test(bmisbp.table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  bmisbp.table
## X-squared = 225.24, df = 3, p-value < 2.2e-16
```

```
prop.trend.test( n.hisbp, n.strata )
```

```
##
##  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
```

Since the p-value < 0.05, we reject the null hypothesis (BMI and SBP are independent) and conclude the alternate (BMI and SBP are not independent). Therefore, BMI is associated with systolic blood pressure (SBP).

Test of independence and test for trend for BMI and CHD,

```
bmichd.table = xtabs( ~ bmigrp + chdfate, data=framingham )
epitab( bmichd.table, pvalue="chi2" )
```

```
## $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"
```

```
n.chd = bmichd.table[,2]
n.strata = rowSums(bmichd.table)
chisq.test(bmichd.table)
```

```
##
##  Pearson's Chi-squared test
##
## data:  bmichd.table
## X-squared = 120.25, df = 3, p-value < 2.2e-16
```

```
prop.trend.test( n.chd, n.strata )
```

```
##
##  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
```

Since the p-value < 0.05, we reject the null hypothesis (BMI and CHD are independent) and conclude the alternate (BMI and CHD are not independent). Therefore, BMI is associated with CHD.

Therefore, the data suggest that your investigators are correct in their prior belief that BMI is a potential cofounder in the relationship between systolic blood pressure and the prevalence of CHD because BMI is associated with both systolic blood pressure and the prevalence of CHD.

4c)

```
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,]
```

Odds ratio and corresponding CI for association between BMI and SBP,

BMI 20-25: OR = 1.62, CI = [1.04, 2.51]

BMI 25-30: OR = 3.21, CI = [2.08, 4.95]

BMI >=30: OR = 6.62, CI = [4.22, 10.4]

Odds ratio and corresponding CI for association between BMI and CHD,

BMI 20-25: OR = 2.16, CI = [1.49, 3.13]

BMI 25-30: OR = 3.65, CI = [2.53, 5.28]

BMI >=30: OR = 4.41, CI = [2.98, 6.53]

4d)

```
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 )
```

The Mantel-Haenszel estimate of the common odds ratio = 1.66 with CI = [1.43, 1.92]

4e)

```
framingham$bmi_categories = cut(framingham$bmi,
  breaks = c(min(framingham$bmi, na.rm=TRUE) - 1, 20, 25, 30, max(framingham$bmi, na.rm=TRUE)+1), i
  labels = c("Low", "Normal", "High", "Obese"))

framingham$sbphi = cut( framingham$sbp, breaks=c(min(framingham$sbp),146, max(framingham$sbp)), inc

ntrt = table(framingham$sbphi, framingham$bmi_categories)[2,]
nctrl = table(framingham$sbphi, framingham$bmi_categories)[1,]
ptrt = table(framingham$sbphi, framingham$bmi_categories, framingham$chdfate)[2,,2]
pctrl = table(framingham$sbphi, framingham$bmi_categories, framingham$chdfate)[1,,2]

bd.test = meta.MH(ntrt, nctrl, ptrt, pctrl)
bd.test
```

```
## Fixed effects ( Mantel-Haenszel ) Meta-Analysis
## Call: meta.MH(ntrt = ntrt, nctrl = nctrl, ptrt = ptrt, pctrl = pctrl)
## Mantel-Haenszel OR =1.66    95% CI ( 1.43, 1.92 )
## Test for heterogeneity: X^2( 3 ) = 4.03 ( p-value 0.2579 )
```

```
summary(bd.test)
```

```
## Fixed effects ( Mantel-Haenszel ) meta-analysis
## Call: meta.MH(ntrt = ntrt, nctrl = nctrl, ptrt = ptrt, pctrl = pctrl)
## -------------------------------------
##        OR (lower  95% upper)
## [1,] 3.78    1.48       9.66
## [2,] 1.84    1.41       2.40
```

```
## [3,] 1.55    1.25       1.92
## [4,] 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 )
```

The null is H0 : OR1 = OR2. Alternate HA: Null is not true.

Since the p-value > 0.05, Breslow-Day test fails to reject the null that all the odds ratios are equal, which opens us up to use the MH test and MH common odds ratio.