Cameron Chandra

10/16/17

Stats449/Green

Homework 4

2.

1. Null hypothesis - H0: theta(xyz1) = theta(xyz2) = 1.

CMH statistic = 16.2282 ; critical value= 3.84

The large CMH statistic and low p-value (see output) allows us to reject the null hypothesis. The relationship between ECG size and coronary disease is not independent given gender. The common odds ratio between X and Y, given Z, does not equal 1.

1. Null hypothesis – H0: theta(xyz1) = theta(xyz2)

Breslow Day statistic = 0.2238; p-value = .6361.

The large p-value fails to reject the null hypothesis (0.6361 > 0.05).

1. Because we not able to reject the null hypothesis of the Breslow Day test, there may be a common odds ratio. As seen in the output in the “Code Used” section, the Mantel-Haenszel estimator is 3.8091 and the 95% confidence interval is (1.9631,7.3911). Because the value 1 is not found in the confidence interval, we can assert that the odds of Coronary Disease is about 38% higher when the ECG size is greater than 0.1.

Code Used:

options nodate pageno=1 formdlim='';

data ECG;

input Gender ECG CD count @@;

datalines;

1 1 1 21 1 1 2 21 1 2 1 9 1 2 2 29

2 1 1 37 2 1 2 10 2 2 1 20 2 2 2 24

;

run;

proc freq data= ECG;

weight count;

tables Gender\*ECG\*CD / cmh;

run;

**The FREQ Procedure**

**Summary Statistics for ECG by CD  
Controlling for Gender**

| **Cochran-Mantel-Haenszel Statistics (Based on Table Scores)** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Statistic** | | **Alternative Hypothesis** | | **DF** | **Value** | | **Prob** | |
| **1** | | **Nonzero Correlation** | | 1 | 16.2282 | | <.0001 | |
| **2** | | **Row Mean Scores Differ** | | 1 | 16.2282 | | <.0001 | |
| **3** | | **General Association** | | 1 | 16.2282 | | <.0001 | |
| **Breslow-Day Test for Homogeneity of the Odds Ratios** | | | |
| **Chi-Square** | | 0.2238 | |
| **DF** | | 1 | |
| **Pr > ChiSq** | | 0.6361 | |
| **Common Odds Ratio and Relative Risks** | | | | | | | |
| **Statistic** | **Method** | | **Value** | **95% Confidence Limits** | | | |
| **Odds Ratio** | Mantel-Haenszel | | 3.8091 | 1.9631 | | 7.3911 | |

3.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Logit Model | Logit Model *N* | Logit Model *p* | Loglinear Model | Loglinear Model *N* | Loglinear Model *p* | *Df* | G^2 |
| α | 4 | 1 | (y,xz) | 8 | 6 | 3 | 27.889 |
| α+βX | 4 | 2 | (xy,xz) | 8 | 6 | 2 | 12.492 |
| α+βZk | 4 | 2 | (xz,yz) | 8 | 6 | 2 | 17.0114 |
| α+βXi+βZk | 4 | 3 | (xy,xz,yz) | 8 | 7 | 1 | 0.2234 |

1. As seen in the section “Code Used”, the residual deviance for each logit model is equivalent to the residual deviance of its corresponding loglinear model. Therefore, these logit and loglinear models are equivalent.

(1) By comparing the residual deviance of the loglinear model (XZ,YZ) and the CMH statistic computed in part 2, we see that the statistics are similar (17.0114 and 16.2282 respectively).

(2) By comparing the residual deviance of the loglinear model (xy,xz,yz) and the Breslow Day statistic computed in part 2, we see that the statistics are similar (0.2234 and 0.2218 respectively).

(3) Yes, you can estimate a common odds ratio. In part 2c, we calculated the common odds ratio for the logit model. The estimate is 3.8091 and the confidence interval is (1.9631,7.3911). For the loglinear model, we must exponentiate the estimate of the interaction between X and Y (exp(ˆβX2)). Therefore, exp(1.22945) = 3.419348. The confidence interval for this estimate is (1.823719, 6.411043). This means that there is a person’s risk for Coronary Disease is 34.1% higher when the electrocardiogram measurement is greater than 0.1 ST segment depression. With 95% confidence, we estimate that the true value is percentage is within the confidence interval above.

c. Based on the interaction between X and Y, we find that the collapsibility condition is satisfied. The confidence interval (1.823719,6.411043) does not contain the value 1, so Y and Z are conditionally independent, given X.

Code used:

Logit:

> yes <- c(21,9,37,20)

> no <- c(21,29,10,24)

> ymat <- cbind(yes,no)

> View(ymat)

> x <- rep(c(1,0), 2)

> z <- gl(2,2)

> m1 <- glm(ymat~z, family= binomial)

> m2 <- glm(ymat~x+z, family= binomial)

> anova(m1,m2)

Resid. Df Resid. Dev Df Deviance

1 2 17.0114

2 1 0.2234 1 16.788

> m3 <- glm(ymat~x,family=binomial)

> summary(m3)

Residual deviance: 12.492 on 2 degrees of freedom

Loglinear:

> z <- gl(2,4)

> x <- gl(2,2,length=8)

> y <- gl(2,1,length=8)

> count <- c(21,21,9,29,37,10,20,24)

> m1 <- glm(count~x+y+z+x\*y+x\*z, family=poisson)

> m2 <- glm(count~x+y+z+x\*z+y\*z, family=poisson)

> m3 <- glm(count~x+y+z+x\*z+y\*z+x\*y, fmamily=poisson)

> summary(m1)

Residual deviance: 12.492 on 2 degrees of freedom

> summary(m2)

Residual deviance: 17.011 on 2 degrees of freedom

> summary(m3)

Residual deviance: 0.22341 on 1 degrees of freedom