Cameron Chandra

10/4/17

Stats449/Green

Homework 2

1. The odds ratio of this 2 x 2 contingency table is 4.2957. For children using night lights, the estimated odds of Myopia equal 0.684. This means that there were 68.4 cases of Myopia found in children using night lights for every 100 responses. The estimated odds of Myopia equal 0.157 for children not using night lights. So, there were 15.7 cases of Myopia found in children not using night lights for every 100 responses. The estimated odds ratio of Myopia found in children using night lights is equal to 4.2957 times the estimated odds for children not using a night light.

The 95% confidence interval of the odds ratio is (2.4190,7.6282). With 95% confidence, we can assert that the real odds ratio is within these parameters. Also, because one is not within the interval, there is a significant difference in contracting Myopia between children using and not using night lights.

1. The relative risk in this 2 x 2 contingency table is 2.964. The sample proportion of Myopia cases was 2.964 times higher for the group of children using night lights

The confidence interval of the relative risk is (1.8668,4.7053). We are 95% confident that the true relative risk is within this interval.

1. There are two tests for statistical independence for 2 x 2 contingency tables: The Likelihood Ratio Test and The Pearson Chi-Square Test. See output page two for test results. Because of the large Likelihood Ratio Chi-Square statistic (28.916>3.834), we reject the null hypothesis that the night lights and Myopia are independent. The Pearson Chi-Square Test is similar, resulting in a large statistic and small p-value, rejecting the null hypothesis that the contracting Myopia is independent of using a night light.

Code used:

Parts (a) and (b):

**proc** **format**;

value ExpFmt **1**= "Night Light Used"

**0**= 'Night Light Not Used';

value RspFmt **1**= "Yes"

**0**= "No";

**run**;

**data** Myopia;

input Exposure Response Count;

label response= "Myopia";

datalines;

0 0 114

0 1 18

1 0 115

1 1 78

;

**proc** **sort** data=Myopia;

by descending Exposure descending Response;

**run**;

**proc** **freq** data = Myopia order = data;

format Exposure ExpFmt. Response RspFmt.;

tables Exposure\*Response / chisq relrisk;

exact pchi or;

weight Count;

title 'Case-Control Study of Night lights';

**run**;

Part (c)

options nodate pageno=**1** formdlim="";

**proc** **format**;

value trtf **1**='Used'

**2**= 'Not Used';

value mif **1**= 'Yes'

**2**= 'No';

**run**;

**data** myopia;

input trt mi count;

datalines;

1 1 78

1 2 115

2 1 18

2 2 114

;

**run**;

**proc** **freq** data= myopia;

weight count;

tables trt\*mi / chisq expected measures;

format trt trtf. mi mif.;

**run**;

1. The definition of relative risk is the ratio of the “success” probabilities for the two groups (black murderers and white murders). Relative risk can only be calculated in a case where there are successes and failures. In this case, Y1 is black victims and Y2 is white victims, so there are no successes or failures to consider.
2. Mean = 20; Standard error = 8.039537

See output sheet “Part A Histogram” for histogram.

1. Mean = 1.141346; Standard error = 0.499628

See output sheet “Part B Histogram” for histogram.

This histogram appears to follow a normal distribution compared to the one in part a.

95% confidence interval for theta = (18.52153,21.47847)

The confidence interval in part b is very different from this one.

Code Used:

> par(mar=rep(2,4))

> data <- rmultinom(1000,size=80,prob=c(30,20,10,20)/80)

> hist(data,col='red',breaks=25,main='Part A Histogram')

> logthfun <- function(thta){log((thta[1]\*thta[4])/(thta[2]\*thta[3]))}

> logtheta <- apply(data,2,logthfun)

> hist(logtheta,col='blue',breaks=25,main='Part B Histogram')

> sd(logtheta)

[1] 0.4996268

> mean(data)

[1] 20

> mean(logtheta)

[1] 1.141346

> sd(data)

[1] 8.039537

> qnorm(.95)\*(sd(data)/sqrt(80))

[1] 1.478473

> left <- mean(data)-((qnorm(.95)\*(sd(data)/sqrt(80))))

> right <- mean(data)+((qnorm(.95)\*(sd(data)/sqrt(80))))

> left

[1] 18.52153

> right

[1] 21.47847

2. The null hypothesis in this test is that obesity and CVD are conditionally independent. Using the Cochran-Mantel-Haenszel test of conditional independence, we reject the null hypothesis (CMH stat = 1.5161 < 1.96). See output sheet for results.
3. In the Breslow-Day test for homogeneity, the null hypothesis is that the odds ratio between obesity and CVD is the same at each age level. Because the Breslow-Day statistic (0.0073) is lower than 1.96, we reject the null hypothesis.
4. Because the association between obesity and CVD seem stable across partial tables, it is reasonable to estimate a common odds ratio. The Mantel-Haenszel common odds ratio in this case is 1.5161. So, people who are obese are 1.5161 times more likely to contract CVD. This is evident through the 95% confidence interval (0.9739,2.3602).

Code Used:

**data** obesity\_cvd;

input age obesity cvd count @@;

datalines;

1 1 1 10

1 1 2 90

1 2 1 35

1 2 2 465

2 1 1 36

2 1 2 164

2 2 1 25

2 2 2 175

;

**run**;

**proc** **freq** data = obesity\_cvd;

weight count;

tables age\*obesity\*cvd / ;

**run**;