Lab Assignment: Inference for Categorical Data

Christian Acosta

July 18, 2016

Create a Word document from this R Markdown file for the following exercises. Submit the R markdown file and resulting Word document via D2L Dropbox.

## Exercise 1

Chondromalatia patellae (CP) is a painful inflammation of the patella (kneecap). It can be diagnosed without error arthroscopically. However, a less invasive diagnostic test called Clarkeâs Sign test can also be used to diagnose CP but is not 100% accurate. The following contingency table contains data for 106 patients who were first diagnosed for CP using the Clarkeâs Sign test and then went on to have arthroscopic surgery to confirm or deny the diagnosis (Doberstein ST, Romeyn RL, Reineke DM , J Athl Train. 2008 Apr-Jun;43(2):190-6.).

|  |  |  |  |
| --- | --- | --- | --- |
|  | Have CP | No CP | Total |
| Positive Clarke's Sign | 9 | 27 | 36 |
| Negative Clarke's Sign | 14 | 56 | 70 |
| Total | 23 | 83 | 106 |

### Part 1a

Conduct a chi-square test to determine if the outcome of the Clarke's Sign test and the presence of CP are associated. Let .

### Answer 1a

clarkTable <- matrix(c(9, 14, 27, 56), nrow=2)  
clarkTable

## [,1] [,2]  
## [1,] 9 27  
## [2,] 14 56

chisq.test(clarkTable)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: clarkTable  
## X-squared = 0.1174, df = 1, p-value = 0.7318

#totalCP <- c(23/106, 83/106)  
#chisq.test(x=positiveClarkeSign, p=totalCP)

Null Hypothesis: Clarke's sign test to detect the presence of CP is independent from the overall rate of CP. Alternative Hypothesis: Clarke's test to detect CP is associated with the rate of CP.

Conclusion: Fail to reject the null hypothesis at alpha = .05.

There is not enough evidence to suggest that the results of Clarke's CP test is associated with the overall rate of CP. (p = 0.7318).

### Part 1b

For the population of all patients who have CP, construct and interpret a 95% confidence interval for the proportion of positive Clarke's Sign tests.

### Answer 1b

totalCPproportion <- 23/106  
prop.test(9, 23, p=totalCPproportion, correct = F)

## Warning in prop.test(9, 23, p = totalCPproportion, correct = F): Chi-  
## squared approximation may be incorrect

##   
## 1-sample proportions test without continuity correction  
##   
## data: 9 out of 23, null probability totalCPproportion  
## X-squared = 4.1138, df = 1, p-value = 0.04253  
## alternative hypothesis: true p is not equal to 0.2169811  
## 95 percent confidence interval:  
## 0.2215762 0.5921448  
## sample estimates:  
## p   
## 0.3913043

The 95% confidence interval estimate for the proportion of positive Clarke's Sign tests in individuals who have diagnosed CP is between 22.16% and 59.21%.

### Part 1c

Compare the proportions of patients without CP who get a positive test to the proportion of patients with CP who get a positive test. Do this by computing a 95% CI for difference of proportions of positive tests of those who have CP and those who don't. Interpret the result.

### Answer 1c

haveCPpositivetest <- c(9, 23)  
noCPpositivetest <- c(27, 83)  
prop.test(haveCPpositivetest, noCPpositivetest, correct = F)

##   
## 2-sample test for equality of proportions without continuity  
## correction  
##   
## data: haveCPpositivetest out of noCPpositivetest  
## X-squared = 0.3122, df = 1, p-value = 0.5763  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.1459835 0.2584333  
## sample estimates:  
## prop 1 prop 2   
## 0.3333333 0.2771084

There is not a statistically significant relationship between those patients without CP who get a positive test result and those patients with CP who get a positive test result.

### Part 1d

If one person is selected randomly from these 106 patients, what is the risk of having CP given that the patient has a positive Clarkeâs Sign test? (âriskâ means probability)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Have CP | No CP | Total |
| Positive Clarke's Sign | 9 | 27 | 36 |
| Negative Clarke's Sign | 14 | 56 | 70 |
| Total | 23 | 83 | 106 |

### Answer 1d

#install.packages("mosaic")  
require("mosaic")

## Loading required package: mosaic

## Warning: package 'mosaic' was built under R version 3.1.3

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.1.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.1.3

## Loading required package: car

## Warning: package 'car' was built under R version 3.1.3

## Loading required package: mosaicData

## Warning: package 'mosaicData' was built under R version 3.1.3

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:car':  
##   
## logit

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cov, D, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

oddsRatio(clarkTable, verbose = T)

##   
## Odds Ratio  
##   
## Proportions  
## Prop. 1: 0.25   
## Prop. 2: 0.2   
## Rel. Risk: 0.8   
##   
## Odds  
## Odds 1: 0.3333   
## Odds 2: 0.25   
## Odds Ratio: 0.75   
##   
## 95 percent confidence interval:  
## 0.3838 < RR < 1.668   
## 0.2886 < OR < 1.949   
## NULL

## [1] 0.75

clarkTable

## [,1] [,2]  
## [1,] 9 27  
## [2,] 14 56

9/36 #The risk of having CP given a positive Clarke's sign test is 0.25 or 25% (9 positive tests w/ confirmed CP of 36 total who tested positive).

## [1] 0.25

### Part 1e

If one person is selected randomly from these 106 patients, what is the risk of having CP given that the patient has a negative Clarkeâs Sign test?

### Answer 1e

14/70 #Relying on the previous code (oddsRatio): The risk of having CP given a negative Clarke's sign test is 0.2 or 20%.

## [1] 0.2

### Part 1f

Compute the relative risk of having CP for those who have a positive Clarke's Sign test compared to those who have a negative test.

### Answer 1f

clarkesSignTest <- oddsRatio(clarkTable, verbose = T)

##   
## Odds Ratio  
##   
## Proportions  
## Prop. 1: 0.25   
## Prop. 2: 0.2   
## Rel. Risk: 0.8   
##   
## Odds  
## Odds 1: 0.3333   
## Odds 2: 0.25   
## Odds Ratio: 0.75   
##   
## 95 percent confidence interval:  
## 0.3838 < RR < 1.668   
## 0.2886 < OR < 1.949   
## NULL

.25/.2

## [1] 1.25

#The relative risk of having CP for those who have a positive Clarke's sign test compared to those who have CP who received a negative Clarke's sign test is 1.25 or 125%.

### Part 1g

What are the odds of having CP given that the patient has a positive Clarkeâs Sign test?

| | Have CP | No CP | Total | | --- | :---: | :---: | :---: | | Positive Clarke's Sign | 9 | 27 | 36 | | Negative Clarke's Sign | 14 | 56 | 70 | Total | 23 | 83 | 106 |

### Answer 1g

9/27

## [1] 0.3333333

#The odds of having CP given the patient has a positive Clarke's Sign test is 0.333 (9 positive w/ CP and 27 positive w/out CP).

### Part 1h

What are the odds of having CP given that the patient has a negative Clarkeâs Sign test?

### Answer 1h

14/56 # The odds of having CP given the patient has a negative Clarke's Sign test is 0.25 (14 positive for CP and 57 negative for CP).

## [1] 0.25

### Part 1i

Compute the odds ratio (OR) of having CP for those who have a positive Clarke's Sign test compared to those who have a negative test.

### Answer 1i

.333333/.25

## [1] 1.333332

#The odds ratio of having CP after a negative Clarke's sign test compared to after a positive test is 1.33.

### Part 1j

Express the odds in the form of a percent.

### Answer 1j

The odds of having CP after a positive Clarke's sign test are 133% of the odds of having CP after a negative Clarke's sign test.

### Part 1k

Construct a 95% confidence interval for the OR of having CP for those who have a positive test compared to those who have a negative test. Interpret the interval, leaving the endpoints as a multiples.

### Answer 1k

#Re-run oddsRatio code.

There is not evidence evidence to suggest a statistically significant relationship at the 95% confidence level.

## Exercise 2

There are said to be six general personality types for dogs (see <http://www.fl-k9.com/personalities.htm> for more details about dog personalities if you are interested). Suppose it has been hypothesized that the distribution of dog personalities is as follows.

|  |  |
| --- | --- |
| Personality Type | Hypothesized Proportion |
| Aggressive | 0.15 |
| Confident | 0.09 |
| Outgoing | 0.29 |
| Adaptable | 0.21 |
| Insecure | 0.12 |
| Independent | 0.14 |

### Part 2a

Consider a random sample of 125 dogs that have been categorized by personality in the table below.

|  |  |
| --- | --- |
| Personality Type | Observed frequency |
| Aggressive | 14 |
| Confident | 15 |
| Outgoing | 22 |
| Adaptable | 30 |
| Insecure | 19 |
| Independent | 25 |
| Total | 125 |

Create vectors for the counts and hypothesized probabilities, then conduct the chi-square test to determine if the distribution of dog personalities given above is correct. Include all parts of the test.

### Answer 2a

observedDogs <- c(14, 15, 22, 30, 19, 15)  
dogPersonalityProportions <- c(.15, .09, .29, .21, .12, .14)  
chisq.test(x=observedDogs, p=dogPersonalityProportions)

##   
## Chi-squared test for given probabilities  
##   
## data: observedDogs  
## X-squared = 10.0158, df = 5, p-value = 0.07479

Null Hypothesis: The population proportions are the same between the random sampling of 125 dogs and hypothesized personality proportions from f1-k9.com.

Alternative Hypothesis: The population proportions are NOT the same between the random sampling of 125 dogs and hypothesized personality proportions from f1-k9.com.

Conclusion: Fail to reject the null hypothesis at alpha = 0.05.

There is insufficient evidence to claim that there is a difference in dog personality types between the random sampling of 125 dogs and the hypothesized dog personality ratios on f1-k9.com (p = 0.075)

### Part 2b

Compute the expected cell counts and verify that they are all large enough for the chi-square test to be valid. Include a reference to the criterion you are using to determine if expected cell counts are large enough.

### Answer 2b

doggyChi <- chisq.test(x=observedDogs, p=dogPersonalityProportions)  
doggyChi$expected

## [1] 17.25 10.35 33.35 24.15 13.80 16.10

The expected sizes of each personality type are larger than the minimum of five recommended for the chi square test. Expected counts for each groups are: Aggressive | 17.25 Confident | 10.35 Outgoing | 33.35 Adaptable | 24.15 Insecure | 13.8 Independent | 16.1

## Exercise 3

A researcher is studying seat belt wearing behavior in teenagers (ages 13 to 19) and senior citizens (over 65). A random sample of 20 teens is taken, of which 17 report always wearing a seat belt. In random sample of 20 senior citizens, 12 report always wearing a seat belt. Using a 5% significance level, determine if seat belt use is associated with these two age groups.

### Part 3a

Create a 2x2 matrix with the correct cell counts. Arrange it so that the columns represent the age group (teen vs senior) and rows contain the seat belt usage (always wear vs not always wear).

### Answer 3a

seatbeltMatrix <- matrix(c(17, 3, 12, 8), nrow=2)  
seatbeltMatrix

## [,1] [,2]  
## [1,] 17 12  
## [2,] 3 8

### Part 3b

With the small cell counts in mind, use the appropriate test to determine if proportions of those who claim to "always wear" a seat belt is the same for these two age groups. Use a 5% significance level. Include all parts for the hypothesis test.

### Answer 3b

#Perform a fisher exact test on the data since cell count is below 5.  
fisher.test(seatbeltMatrix)

##   
## Fisher's Exact Test for Count Data  
##   
## data: seatbeltMatrix  
## p-value = 0.1552  
## alternative hypothesis: true odds ratio is not equal to 1  
## 95 percent confidence interval:  
## 0.6908715 25.8641435  
## sample estimates:  
## odds ratio   
## 3.651794

Null hypothesis: The proportion of seatbelt wearing teens is equal to the proportion of seatbelt wearing seniors. Alternative hypothesis: The proportion of seatbelt wearing teens is NOT equal to the proportion of seatbelt wearing seniors.

Conclusion: We fail to reject the null hypothesis at alpha=0.05.

There is insufficient evidence to suggest that the ratio of seatbelt wearing teens is different than the ratio of seatbelt wearing seniors (p = 0.1552).