Inference for Categorical Data

Your name here

mm/dd/yyyy

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

## Exercise 1

Suppose independent, random samples of credit unions and banks had the following frequencies for being rated as Outstanding.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Outstanding | Not Outstanding |  |
| Bank | 70 | 150 |  |
| Credit Union | 66 | 81 |  |

### Part 1a

Create the table in R from the data and display it with the margins. Include the names for the rows and columns.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 1b

For the population of all credit unions, construct and interpret a 95% confidence interval for the proportion of credit unions rated as Outstanding.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1c

Compare the proportions of credit unions that are rated as Outstanding to the proportion of banks that are rated as Outstanding. Do this by computing a 95% CI for difference in proportions of those rated as Outstanding for credit unions minus banks. Interpret the result.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1d

If one bank is selected randomly, what is the estimated risk of not being rated as Outstanding? (“risk” means probability)

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1e

If one credit union is selected randomly, what is the estimated risk of not being rated as Outstanding?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1f

Compute the relative risk of not being rated as Outstanding for banks compared to credit unions.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1f -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1g

What are the estimated odds of a credit union being rated as Outstanding?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1h

What are the estimated odds of a bank being rated as Outstanding?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1h -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1i

Compute the estimated odds ratio of being rated as Outstanding for credit unions compared to banks.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1i -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1j

Write an interpretation of the odds ratio of being rated as Outstanding for credit unions compared to banks as a percent.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1j -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 1k

Construct a 95% confidence interval for the population odds ratio of being rated as Outstanding for credit unions compared to banks. Interpret the interval, leaving the endpoints as a multiples.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1k -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 1l

Based on the 95% CI for the odds ratio, is there significant evidence of an association between being rated as Outstanding and whether or not an institution is a bank or credit union? Explain.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1l -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

## Exercise 2

Marketing Research reported results of a study of online purchases where demographic information was collected on customers. The age group of the customer (under 18, 18 to 35, 36 to 50, or over 50) purchased by each of 165 consumers was recorded.

### Part 2a

A leading internet market research company claims that 13% of all online purchases are made by customers under 18, 32% by customers between 18 and 35, 38% by customers between 36 and 50, and the remaining 17% by customers over 50 years of age.

Test this claim if sample data shows that 28 customers in the sample were under 18, 44 were 18 to 35, 54 were 36 to 50, and 39 were over 50.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age Group | Under 18 | 18 to 35 | 36 to 50 | Over 50 |
| Count | 28 | 44 | 54 | 39 |

Use . State the hypotheses for the test and use R to compute the test statistic, df, and P-value. State the conclusion, including a practical interpretation in the context of the problem. Include the P-value in the conclusion.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 2c

Display the data in a bar graph and comment on its features.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

## 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 19 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 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

## Exercise 4

A study was conducted whereby the type of anesthetic (A or B), nausea after the surgery (Yes or No), the amount of pain medication taken during the recovery period, and age for a random sample of 72 patients undergoing reconstructive knee surgery.

The data frame is the anesthesia in the DS705data package.

### Part 4a

Create and display a contingency table with the type of anesthetic defining the rows and nausea after the surgery as the columns. Display the margins for this table as well.

Also make a side-by-side bar graph showing the nausea (Yes vs No) on the horizontal axis and color-coded bars to indicate the type of anesthetic.

Comment on any potential relationships between nausea and type of anesthetic you see in the graph.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 4b

Conduct a chi-square test for independence at the 10% level.

State the hypotheses (in words) and conclusion of the test. Use R to compute the test statistic, degrees of freedom, and P-value. Include the P-value in your written conclusion.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 4b -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.