Problem 2.21

Chad Huntebrinker

2025-02-23

Problem 2.21, we are given a 4x4 matrix with the income (in thousands of dollars) and job satisfaction (in different ranges) for black Americans.

For part A, we need to test the independence of job satisfaction and income using X^2 and interpret and explain the deficiency of this test for these data. We also need to find the standardized residuals and see if they suggest any association pattern.

##   
## Pearson's Chi-squared test  
##   
## data: data\_table  
## X-squared = 11.524, df = 9, p-value = 0.2415

## Very Dissatisfied A Little Satisfied Moderately Satisfied Very Satisfied  
## <5 1.4406207 0.7305280 -0.1606257 -1.079180  
## 5-15 0.7525419 0.8715747 0.6005081 -1.772568  
## 15-25 -1.1171392 -1.5211385 0.2198088 1.509794  
## >25 -1.1171392 -0.1573592 -0.7326960 1.509794

What we find is the X^2 is about 11.5. For deficiency, the chi-square test assumes that we have large samples to utilize it. However, if we do not, then the test can have unreliable results. A good way to check this is if any of the expected frequency are less than 5. When we do this, we find a couple of examples (any of the row totals multipled with column 1’s total) has an expected frequency less than 5. This leads to the conclusion that we should probably use a small-sample test for this data.

We also see the standardized residuals suggest that as employees get paid more, their satisfaction grows as well.

For part B, we need to conduct a test that treats the variables in a quantitive manner and explain why the results differ so much.

## M^2: 7.044859

## df: 1

## P-score: 0.007949308

The reason why the results differ so much above is because the chi-square test does not take into account ordered categories, it just looks at if there is an association. The M^2 test statistic, on the other hand, does.

#Code Appenix  
#Chad Huntebrinker  
#Problem 1  
#Part a  
  
data\_table <- matrix(c(2, 4, 13, 3,  
 2, 6, 22, 4,  
 0, 1, 15, 8,  
 0, 3, 13, 8), nrow = 4, byrow = TRUE)  
rownames(data\_table) <- c("<5", "5-15", "15-25", ">25")  
colnames(data\_table) <- c("Very Dissatisfied", "A Little Satisfied", "Moderately Satisfied", "Very Satisfied")  
  
chi\_results <- chisq.test(data\_table)

## Warning in chisq.test(data\_table): Chi-squared approximation may be incorrect

chi\_results

##   
## Pearson's Chi-squared test  
##   
## data: data\_table  
## X-squared = 11.524, df = 9, p-value = 0.2415

chi\_results$stdres

## Very Dissatisfied A Little Satisfied Moderately Satisfied Very Satisfied  
## <5 1.4406207 0.7305280 -0.1606257 -1.079180  
## 5-15 0.7525419 0.8715747 0.6005081 -1.772568  
## 15-25 -1.1171392 -1.5211385 0.2198088 1.509794  
## >25 -1.1171392 -0.1573592 -0.7326960 1.509794

#The standardized residuals suggest that as employees get paid more, their satisfaction grows as well.  
  
#Part B  
income\_score <- c(3, 10, 20, 35)  
sat\_score <- c(1, 3, 4, 5)  
  
cmh\_results <- CMHtest(data\_table, rscores = income\_score, cscores = sat\_score)  
cat("M^2: ", cmh\_results$table[1], "\n") #M^2

## M^2: 7.044859

cat("df: ", cmh\_results$table[5], "\n") #df

## df: 1

cat("P-score: ", cmh\_results$table[9], "\n") #P-score

## P-score: 0.007949308

#The chi-square test does not take into account ordered categories, it just looks at if there is  
#an association. The M^2 test statistic does.