Homework Problem 2

Chad Huntebrinker

2025-02-17

#Chad Huntebrinker  
#Problem 2.17  
  
#Part a  
#Create the original matrix  
race\_gap <- matrix(c(871, 821, 336, 347, 42, 83), ncol = 3, byrow=TRUE)  
  
#Assign row and col names  
rownames(race\_gap) <- c("White", "Black")  
colnames(race\_gap) <- c("Democrat", "Republican", "Independent")  
  
#Calculate the sums for LRS  
sum\_of\_rows <- rowSums(race\_gap)  
sum\_of\_cols <- colSums(race\_gap)  
total <- sum(race\_gap)  
  
#Get the expected frequency matrix  
expected\_frequency <- matrix(c(sum\_of\_rows[1] \* sum\_of\_cols[1] / total, sum\_of\_rows[1] \* sum\_of\_cols[2] / total,  
 sum\_of\_rows[1] \* sum\_of\_cols[3] / total, sum\_of\_rows[2] \* sum\_of\_cols[1] / total,  
 sum\_of\_rows[2] \* sum\_of\_cols[2] / total, sum\_of\_rows[2] \* sum\_of\_cols[3] / total),  
 ncol = 3, byrow=TRUE)  
  
#Calculate the LRS  
LRS <- 2 \* sum(race\_gap \* log(race\_gap / expected\_frequency))  
print(paste("Likelihood-Ratio Statistic: ", LRS))

## [1] "Likelihood-Ratio Statistic: 213.901067936876"

#Get X^2 and p-value  
chi\_test <- chisq.test(race\_gap)  
  
chi\_test

##   
## Pearson's Chi-squared test  
##   
## data: race\_gap  
## X-squared = 184.32, df = 2, p-value < 2.2e-16

chi\_test$p.value

## [1] 9.435104e-41

#Because the p-value is 9.435104e-41, we can interpret that H0 is invalid. As result,  
#we can say there is no independence and there is an association between race and political party.  
  
#Part b  
#Get the standardized residuals  
chi\_test$stdres

## Democrat Republican Independent  
## White -11.96679 12.99946 -0.5326281  
## Black 11.96679 -12.99946 0.5326281

#The standardized residuals for White Democrats and Black Republicans show strong evidence of fewer  
#people in these cells than if they were independent of race. And the standardized residuals of White  
#Republican and Black Democrat show strong evidence of more people in these cells than if they were  
#independent of race.  
  
#Part c  
matrix\_1 <- matrix(c(871, 821, 347, 42), ncol = 2, byrow=TRUE)  
  
sum\_of\_rows <- rowSums(matrix\_1)  
sum\_of\_cols <- colSums(matrix\_1)  
total <- sum(matrix\_1)  
  
#Calculate the expected frequency matrix  
expected\_frequency <- matrix(c(sum\_of\_rows[1] \* sum\_of\_cols[1] / total, sum\_of\_rows[1] \* sum\_of\_cols[2] / total,  
 sum\_of\_rows[2] \* sum\_of\_cols[1] / total, sum\_of\_rows[2] \* sum\_of\_cols[2] / total),  
 ncol = 2, byrow=TRUE)  
  
#Calculate the LRS  
LRS2 <- 2 \* sum(matrix\_1 \* log(matrix\_1 / expected\_frequency))  
print(paste("Likelihood-Ratio Statistic: ", LRS2))

## [1] "Likelihood-Ratio Statistic: 213.620150014379"

LRS - LRS2

## [1] 0.2809179

pchisq(LRS2, 1, lower.tail = FALSE)

## [1] 2.228861e-48

pchisq(LRS - LRS2, 1, lower.tail = FALSE)

## [1] 0.5961002

#LRS2 is for comparing races on (Democrat, Republican) choice and the above score of about 0.3  
#is for comparing races on (Democrat + Republican, Independent). According to their p-score, there is  
#strong evidence that of a difference in the relative numbers between races and  
#if they identify as Democrat or Republican but no strong evidence between races in the relative  
#numbers identifying as Independent instead of Democrat or Republican.