Assignment 6

Aishat Olatunji

2022-09-29

data = read.csv("/Users/aishatolatunji/Downloads/Americandata.csv")  
head(data)

## Happy Students Laborers Preachers Physicians Housewives Teachers Lawyers  
## 1 Yes 390 378 35 159 78 108 11  
## 2 No 1610 122 265 51 122 38 64  
## Musicians  
## 1 31  
## 2 19

dim(data)

## [1] 2 9

# removing the categorical colum i.e the first column  
data=data[-1]  
head(data)

## Students Laborers Preachers Physicians Housewives Teachers Lawyers Musicians  
## 1 390 378 35 159 78 108 11 31  
## 2 1610 122 265 51 122 38 64 19

dim(data)

## [1] 2 8

#conversion to matrix form  
data\_matrix<- data.matrix(data)  
rownames(data\_matrix)<-c("yes","No")  
data\_matrix

## Students Laborers Preachers Physicians Housewives Teachers Lawyers  
## yes 390 378 35 159 78 108 11  
## No 1610 122 265 51 122 38 64  
## Musicians  
## yes 31  
## No 19

# obtaining the chi square  
  
library(gmodels)  
CrossTable(data\_matrix,chisq=TRUE,expected=TRUE,sresid=TRUE,format="SPSS")

##   
## Cell Contents  
## |-------------------------|  
## | Count |  
## | Expected Values |  
## | Chi-square contribution |  
## | Row Percent |  
## | Column Percent |  
## | Total Percent |  
## | Std Residual |  
## |-------------------------|  
##   
## Total Observations in Table: 3481   
##   
## |   
## | Students | Laborers | Preachers | Physicians | Housewives | Teachers | Lawyers | Musicians | Row Total |   
## -------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|  
## yes | 390 | 378 | 35 | 159 | 78 | 108 | 11 | 31 | 1190 |   
## | 683.712 | 170.928 | 102.557 | 71.790 | 68.371 | 49.911 | 25.639 | 17.093 | |   
## | 126.174 | 250.859 | 44.501 | 105.943 | 1.356 | 67.607 | 8.359 | 11.315 | |   
## | 32.773% | 31.765% | 2.941% | 13.361% | 6.555% | 9.076% | 0.924% | 2.605% | 34.186% |   
## | 19.500% | 75.600% | 11.667% | 75.714% | 39.000% | 73.973% | 14.667% | 62.000% | |   
## | 11.204% | 10.859% | 1.005% | 4.568% | 2.241% | 3.103% | 0.316% | 0.891% | |   
## | -11.233 | 15.839 | -6.671 | 10.293 | 1.164 | 8.222 | -2.891 | 3.364 | |   
## -------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|  
## No | 1610 | 122 | 265 | 51 | 122 | 38 | 64 | 19 | 2291 |   
## | 1316.288 | 329.072 | 197.443 | 138.210 | 131.629 | 96.089 | 49.361 | 32.907 | |   
## | 65.538 | 130.302 | 23.115 | 55.029 | 0.704 | 35.117 | 4.342 | 5.877 | |   
## | 70.275% | 5.325% | 11.567% | 2.226% | 5.325% | 1.659% | 2.794% | 0.829% | 65.814% |   
## | 80.500% | 24.400% | 88.333% | 24.286% | 61.000% | 26.027% | 85.333% | 38.000% | |   
## | 46.251% | 3.505% | 7.613% | 1.465% | 3.505% | 1.092% | 1.839% | 0.546% | |   
## | 8.096 | -11.415 | 4.808 | -7.418 | -0.839 | -5.926 | 2.084 | -2.424 | |   
## -------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|  
## Column Total | 2000 | 500 | 300 | 210 | 200 | 146 | 75 | 50 | 3481 |   
## | 57.455% | 14.364% | 8.618% | 6.033% | 5.745% | 4.194% | 2.155% | 1.436% | |   
## -------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|  
##   
##   
## Statistics for All Table Factors  
##   
##   
## Pearson's Chi-squared test   
## ------------------------------------------------------------  
## Chi^2 = 936.1395 d.f. = 7 p = 7.523435e-198   
##   
##   
##   
## Minimum expected frequency: 17.09279

## Reports  
  
# Chi^2 = 936.1395 d.f. = 7 p = 7.523435e-198   
# Minimum expected frequency: 17.09279   
   
# chi-squared = 936.1395   
# degrees of freedom = (8-1) = 7  
# p. value = 7.523435e-198   
  
# hypothesis testing  
# Ho : The type of profession related to the happiness of american people is not statistically significant  
  
# H1 : The type of profession related to the happiness of american people is statistically significant  
  
# Decision rule: since the pvalue(7.523435e-198 ) < alpha we reject the null hypothesis (Ho) and conclude that the model is statistically significant.that is; the type of profession american people does is related to their happiness.  
  
## The Minimum expected frequency is 17.09279 which shows that the expected frequency is greater than 5, therefore it shows that it follows the assumption of independence.  
  
  
## explaining the chi\_squared   
# since the chi-squared (936.1395) > critical region at 8degrees of freedom (15.507) we reject the null hypothesis and conclude that The chi- square also shows that they are the type of profession american people does is related to their happiness.  
  
   
  
  
# optional step  
  
# For the housewives based on the chi square contribution of 1.356 it shows that The type of profession related to the happiness of american people is random i.e independent.  
  
# ------- Final optional step  
  
# standard residual are referred to as a standard score, they are often defined mathematically as the differnce between observed and expected divided by the square root of an expected value and they are often used to determine which attributes contributes more to the significance of a chi square e.g lets use the happy housewives as an example (78-68.371)/sqrt(68.371) = 1.164515 which from the output here shows that it isnt a major contributor.