HW6

Amane Chibana and Harry Wang

2024-03-21

Problem 1

One categorical random variable is the version of the product viewed. Which can have four possible outcomes, where either the individual views the first, second, third or fourth version of the product

The second categorical random variable is the willingness to buy the product. This has two possible outcomes, where the individual is likely to buy it, or unlikely to buy.

Problem 2

```
#T(row number)(column number)
T11 = 25
T12 = 18
T13 = 30
T14 = 24
T21 = 35
T22 = 17
T23 = 15
T24 = 26

observed = matrix(c(T11,T12,T13,T14,T21,T22,T23,T24),nrow=4)

rownames(observed) = c('1st', '2nd', '3rd', '4th')
colnames(observed) = c('Likely to buy', 'Unlikely to buy')

observed=as.table(observed);observed
```

```
##     Likely to buy Unlikely to buy
## 1st            25              35
## 2nd            18              17
## 3rd            30              15
## 4th            24              26
```

Problem 3

```
#Totals
Likely = 25+18+30+24
Unlikely = 190-Likely
```

```r
first = 25+35
second = 18+17
third = 30+15
fourth = 24+26
total=190

#T(row number)(column number)
#(Total of Col)*(Total of Row/Total)
T11 = round(Likely*(first/total),digits = 2)
T12 = round(Likely*(second/total),digits = 2)
T13 = round(Likely*(third/total),digits = 2)
T14 = round(Likely*(fourth/total),digits = 2)
T21 = round(Unlikely*(first/total),digits = 2)
T22 = round(Unlikely*(second/total),digits = 2)
T23 = round(Unlikely*(third/total),digits = 2)
T24 = round(Unlikely*(fourth/total),digits = 2)
T11;T12;T13;T14;T21;T22;T23;T24
```

```
## [1] 30.63
```

```
## [1] 17.87
```

```
## [1] 22.97
```

```
## [1] 25.53
```

```
## [1] 29.37
```

```
## [1] 17.13
```

```
## [1] 22.03
```

```
## [1] 24.47
```

```r
expected = matrix(c(T11,T12,T13,T14,T21,T22,T23,T24),nrow=4)

rownames(expected) = c('1st', '2nd', '3rd', '4th')
colnames(expected) = c('Likely to buy', 'Unlikely to buy')

expected=as.table(expected);expected
```

```
##     Likely to buy Unlikely to buy
## 1st         30.63           29.37
## 2nd         17.87           17.13
## 3rd         22.97           22.03
## 4th         25.53           24.47
```

Problem 4

```r
#Observed
OT11 = 25
OT12 = 18
OT13 = 30
OT14 = 24
OT21 = 35
OT22 = 17
OT23 = 15
OT24 = 26

#Totals
Likely = 25+18+30+24
Unlikely = 190-Likely
first = 25+35
second = 18+17
third = 30+15
fourth = 24+26
total=190

#T(row number)(column number)
#(Total of Col)*(Total of Row/Total)
T11 = round(Likely*(first/total),digits = 2)
T12 = round(Likely*(second/total),digits = 2)
T13 = round(Likely*(third/total),digits = 2)
T14 = round(Likely*(fourth/total),digits = 2)
T21 = round(Unlikely*(first/total),digits = 2)
T22 = round(Unlikely*(second/total),digits = 2)
T23 = round(Unlikely*(third/total),digits = 2)
T24 = round(Unlikely*(fourth/total),digits = 2)

teststat= ((OT11-T11)^2)/T11 +((OT12-T12)^2)/T12+((OT13-T13)^2)/T13 +((OT14-T14)^2)/T14 +((OT21-T21)^2)/
```

```
## [1] 6.698234
```

```r
row=4
col=2
df = (row-1)*(col-1);df
```

```
## [1] 3
```

The X^2 test statistic is 6.70 with 3 degree of freedom.

Problem 5

(i)

H_0 : There is probably no association between categorical variables, version viewed and likeliness of purchase.

H_a : There is a possible association between categorical variables, version viewed and likeliness of purchase.

(ii)

The null distribution of the chi-squared test statistic is a chi-squared distribution with degrees of freedom (r-1)(c-1) = (4-1)(2-1) = 3. So the null distribution is chi-squared with 3 degrees of freedom.

Problem 6

```
chi_square_statistic <- 6.7

df = 3

p_value = 1 - pchisq(chi_square_statistic, df)

print(paste("P-value:", round(p_value,4)))
```

```
## [1] "P-value: 0.0821"
```

Problem 7

Given the calculated p-value of 0.0821 and considering a common significance level : 0.05

Since the p-value (0.0821) is greater than the significance level (0.05), we fail to reject the null hypothesis. This means there is not enough statistical evidence at the 5% significance level to conclude that there is an association between the version of the product and the willingness of potential customers to buy the product.

Problem 8

Failing to reject the null hypothesis indicates there's no significant evidence that different versions of the product affect customer willingness to buy. This suggests that variations in purchase intent among the product versions could be due to chance. It implies that factors other than product version, like marketing or price, might be more influential on purchase decisions. The company may need to focus on these aspects to enhance customer willingness to buy.

Problem 9

```
# Performing the Chi-squared test
test_result <- chisq.test(observed)

# Printing the test result
print(test_result)
```

```
##
##  Pearson's Chi-squared test
##
## data:  observed
## X-squared = 6.694, df = 3, p-value = 0.08232
```

The p-value is greater than 0.05 , we fail to reject the null hypothesis, indicating that there is not enough evidence to suggest a significant association between the product version and customer willingness to buy.