Faculty of Computing
Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics Lab Sheet 10

Lab Exercise 10 - Chi Squared Tests Week 13

Before starting the lab sheet, you need to create a folder in your desktop and save all your working inside the folder. Set the working directory to that folder using the following command:

setwd("paste the path of the folder")

Eg:- setwd("D:\\2025 - Sem 2\\IT2120\\Lab Sessions\\Lab 10")

1. A shop owner claims that an equal number of customers come into his shop each weekday. To test this hypothesis, a researcher records the number of customers that come into the shop in a given week and finds the following:

Monday: 55 customers
Tuesday: 62 customers
Wednesday: 43 customers
Thursday: 46 customers
Friday: 50 customers

i. Conduct a suitable Chi-square test to check the claim of shop owner.

```
setwd("D:\\2025 - Sem 2\\IT2120 - New\\Lab Sessions\\Lab 10")

##Question 01
#Part 1
#Here, the shop owner claims that an equal number of customers come into his shop each weekday.
#That means probability of customer arriving on each day would be 0.2.
#To test this claim we need to conduct goodness of fit test which is a chi-squared test.
#So that null hypothesis will be probability that customers arriving on each day will be 0.2.
#Alternative hypothesis will be at least one weekday exists such that probability of customer
#arriving will be different from 0.2.
#To conduct the test observed counts will be stored into a variable called "observed"
#And probabilities for each day will be stored into another variable called "prob"
observed <- c(55, 62, 43, 46, 50)
prob <- c(.2, .2, .2, .2, .2)
#To conduct the test "chisq.test" command will be used as follows.
chisq.test(x=observed, p=prob)
```

ii. Write your conclusion based on test results.

```
#Part 2
#Consider 5% level of significance for the test.
#Rejection Region: If the p value for the test is less than 0.05,
#reject the null hypothesis at 5% level of significance.
#P value for the test got as 0.351.
#Conclusion:Since the p value (0.351) is greater than 0.05, do not reject null hypothesis at 5%
#level of significance. Therefore we can conclude that probability that customers arriving on
#each day will be same which is 0.2.
```

2. Consider the house tasks data set in the path, http://www.sthda.com/sthda/RDoc/data/housetasks.txt which contains contingency table with 13 house tasks and their distribution in the couple. An image of the data is displayed below:

	Wife	Alternating	Husband	Jointly
Laundry	156	14	2	4
Main_meal	124	20	5	4
Dinner	77	11	7	13
Breakfeast	82	36	15	7
Tidying	53	11	1	57
Dishes	32	24	4	53
Shopping	33	23	9	55
Official	12	46	23	15
Driving	10	51	75	3
Finances	13	13	21	66
Insurance	8	1	53	77
Repairs	0	3	160	2
Holidays	0	1	6	153

Here rows are the different tasks, values are the frequencies of the tasks done: by the wife only, alternatively, by the husband only or jointly.

i. Import the data into R.

```
##Question 02
#Part 1
#First needs to set the path where data set exists as follows.
file_path <- "http://www.sthda.com/sthda/RDoc/data/housetasks.txt"
#Then,"read.delim" command will be used to reads the file which is in a table format
#and creates a data frame from it, with cases corresponding to lines and variables to fields.
housetasks <- read.delim(file_path, row.names = 1)
housetasks
```

ii. Test whether there is any association between house tasks and their distribution in the couple.

```
#Part 2
#Here, they wanted to check whether there is any association between house tasks
#and their distribution in the couple.
#To test this we need to conduct the association test which is a chi-squared test.
#So that null hypothesis will be there is no association between house tasks
#and their distribution in the couple.
#Alternative hypothesis will be there is a significant association between house tasks
#and their distribution in the couple.
#To conduct the test "chisq.test" command will be used as follows.
chisq <- chisq.test(housetasks)
chisq
#Consider 5% level of significance for the test.
#Rejection Region: If the p value for the test is less than 0.05,
#reject the null hypothesis at 5% level of significance.
#P value for the test got as 2.2e-16.
#Conclusion:Since the p value (2.2e-16) is less than 0.05, reject the null hypothesis at 5%
#level of significance. Therefore we can conclude that there is a significant association between
#house tasks and their distribution in the couple.
```

Exercise

Instructions: Create a folder in your desktop with your registration number (Eg: "IT....."). You need to save the R script file and take screenshots of the command prompt with answers and save it in a word document inside the folder. Save both R script file and word document with your registration number (Eg: "IT....."). After you finish the exercise, zip the folder and upload the zip file to the submission link.

1. A vending machine owner claims that customers choose the four snack types (A, B, C, D) with equal probability. To test this claim, a researcher records the number of purchases for each snack type during one week and results are given below.

Snack_Type	Count
A	120
B	95
C	85
D	100

   i. State the null and alternative hypotheses for the test.

   ii. Perform a suitable chi-squared test to test the null hypothesis.

   iii. Give your conclusions based on the results.

```
Console   Terminal   Background Jobs
R 4.5.1 · C:/Users/Geenuth/2nd Year 1st Semester/Probability and Statistics(IT2120)/Practical/Lab 10/
> setwd("C://Users//Geenuth//2nd Year 1st Semester//Probability and Statistics(IT2120)//Practical//
Lab 10")
> getwd()
[1] "C:/Users/Geenuth/2nd Year 1st Semester/Probability and Statistics(IT2120)/Practical/Lab 10"
>
> #Exercise
> #Question 01
> snack_counts <- c(120, 95, 85, 100)
> chisq.test(snack_counts, p = rep(1/4, 4))

        Chi-squared test for given probabilities

data:  snack_counts
X-squared = 6.5, df = 3, p-value = 0.08966

>
> #i.State the null and alternative hypotheses for the test.
> #H₀ (Null Hypothesis): Customers choose each snack type (A, B, C, D) with equal probability.
> #H₁ (Alternative Hypothesis): Customers do not choose each snack type equally.
>
> #ii.Perform a suitable chi-squared test to test the null hypothesis.
> chisq.test(snack_counts, p = rep(1/4, 4))

        Chi-squared test for given probabilities

data:  snack_counts
X-squared = 6.5, df = 3, p-value = 0.08966
```

```
>
> #iii.Give your conclusions based on the results
> #If p-value > 0.05, fail to reject H₀ → the data supports that customers choose equally.
> #If p-value < 0.05, reject H₀ → the data shows that customers don't choose equally.
>
```