

MINI PROJECT

A STATISTICAL PROJECT ON THE INVOLVEMENT OF YOUTH IN CYBERCRIME IN INDIA

*Submitted to Department of Mathematics and
Statistics*

**SACRED HEART COLLEGE (AUTONOMOUS),
THEVARA, KOCHI**

SUBMITTED BY,

ROHITH JACOB ABRAHAM

MEGHNA BABU

CERTIFICATE

*This is to certify that the mini project entitled “**INVOLVEMENT OF YOUTH IN CYBERCRIME IN INDIA**” is a bonafide work carried out by **Meghna Babu and Rohith Jacob Abraham** during the year **2022 – 23** under the guidance of **Asst. Prof. Dr. Lakshmipriya R** , Department of Statistics & Mathematics, Sacred Heart College (AUTONOMOUS), Thevara, Kochi.*

Dr. Lakshmipriya R
Assistant Professor
Department of Statistics
& Mathematics.
Sacred Heart College
(AUTONOMOUS)

ACKNOWLEDGEMENT

First of all, we would like to thank god almighty for giving us the opportunity and guidance to achieving our goal and to be successful in this part.

We take this opportunity to express our sincere and heartfelt gratitude to our research supervisor Dr. Lakshmipriya R, Assistant Professor , Department of Mathematics And Statistics, Sacred Heart Collage, Thevara, for her constant guidance and regular interaction throughout our research work.

We express our deep sense of gratitude to Dr. Jeenu, Head of Department of Mathematics, for providing the facilities to accomplish the research work and for being an inspiration for the successful completion of our project.

We also thank our friends and family for their constant encouragement and support in various phases of the completion of the project.

CONTENTS

1. INTRODUCTION

2. OBJECTIVE

3. METHODOLGY

4. RESULT AND DISCUSSIONS

5. DATA ANALYSIS

6. CONCLUSION

7. REFERENCES

1) INTRODUCTION

Usage of Internet has become a daily routine for majority of people for day-to-day transactions.

With the evolution of technology as we step into the digital world, cybercrime is now on the rise. With each step, new challenges and responsibilities arise day by day. The population focused on aggression is young people and adolescents. They are the major population in the world enjoying digitization. Cybercrime is a major threat to a growing society.

Cybercrime has a huge impact on young people. Young people are more likely to access the Internet than any other demographic because they are the first victims of cybercrime. Themes like fitness, especially emotional well-being, are meant to explain many cybercrimes. Many young people are trying to commit suicide because they are victims of cybercrime. This paper will examine all the consequences of cybercrime, including cyber bullying (a type of cybercrime), and how young people can prevent cybercrime.

In short, this unique combination of digital gadgets such as smart phones and the Internet allows us to instantly communicate with others, family and friends around the world.

We depends heavily on information provided by the Internet for office work, e-commerce, banking, weather forecasting, business deals, fitness tips, stock markets, entertainment, recreation, satisfying mental motivations and emotions, and pass-time activity. Upload, share, download, Google IT, apps, etc. are just some of the common expressions that work at your fingertips. So it is no exaggeration to say that smart phones and other internet-enabled personal electronic gadgets have penetrated all walks of life and opened the door to cybercrime. Lack of awareness on such issues can lead to serious damage for financial, emotional, moral or ethical reasons. In such a dire situation, in addition to dealing with cybercrime, there is another issue that should be given high priority - to create awareness among Internet users about "cybercrime and security".

2) OBJECTIVE

- 1. The study highlights the trends in cybercrime and the steps that can be taken to address this rapidly growing problem.**
- 2. To know the various factors which lead the youngsters into cybercrimes.**
- 3. Through this survey we give students an idea of how to protect yourself from cybercrime.**
- 4. An understanding of the emerging cyber threat and knowledge of how cyber response strategy can be an integral part of the overall framework is important.**
- 5. We conduct research to circulate information and raise awareness about cyber security and cybercrime.**

3) **METHODOLOGY**

The current study is based on quantitative research analysis. A questionnaire was prepared to include questions aimed at young people in various walks of life to understand the nature and type of cybercrime prevalent in society, how to deal with it, and the awareness, impact and challenges of cybercrime. Then an online survey was conducted.

From April 11 to April 14, 2022, the Google form link containing the questionnaire was shared among various youths on various social media platforms and asked the youth to complete the questionnaire. From the survey, we received a random sample of 69 young people of different ages, including 50 girls and 19 boys.

After receiving the data, the data was analyzed with the help of R software and Microsoft Excel.

STATISTICAL TOOLS AND MEASURES USED

ARITHMETIC MEAN

The arithmetic mean is the simplest and most widely used measure of a mean, or average. It simply involves taking the sum of a group of numbers, then dividing that sum by the count of the numbers used in the series.

Formula

$$A = \frac{1}{n} \sum_{i=1}^n a_i$$

A = arithmetic mean

n = number of values

a_i = data set values

STANDARD DEVIATION

Standard Deviation is a measure which shows how much variation (such as spread, dispersion, spread,) from the mean exists. The standard deviation indicates a “typical” deviation from the mean. It is a popular measure of variability because it returns to the original units of measure of the data set. Like the variance, if the data points are close to the mean, there is a small variation whereas the data points are highly spread out from the mean, then it has a high variance. Standard deviation calculates the extent to which the values differ from the average. Standard Deviation, the most widely used measure of dispersion, is based on all values. Therefore a change in even one value affects the value of standard deviation. It is independent of origin but not of scale. It is also useful in certain advanced statistical problems.

<u>Sample</u>	<u>Population</u>
$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$	$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$

GRAPHS AND CHARTS (BAR, PIE, DOUGHNUT)

Graph or Chart a statistical graph or chart is defined as the pictorial representation of statistical data in graphical form. The statistical graphs are used to represent a set of data to make it easier to understand and interpret statistical information. The basic graphs used in statistics include bar, line, histogram and pie charts. These are explained here in brief.

Bar Graph

Bar graphs are the pictorial representation of grouped data in vertical or horizontal rectangular bars, where the length of bars is proportional to the measure of data. The chart's horizontal axis represents categorical data, whereas the chart's vertical axis defines discrete data.

Pie Chart

A pie chart used to represent the numerical proportions of a dataset. This graph involves dividing a circle into various sectors, where each sector represents the proportion of a particular element as a whole. This is also called a circle chart or circle graph.

Doughnut chart

Doughnut chart is a type of chart in excel whose function of visualization is just similar to pie charts, the categories represented in this chart are parts and together they represent the whole data in the chart, only the data which are in rows or columns only can be used in creating a doughnut chart in excel, however it is advised to use this chart when we have less number of categories of data.

CHI- SQUARE TEST FOR INDEPENDENCE

The Chi-Square test is a statistical procedure for determining the difference between observed and expected data. This test can also be used to determine whether it correlates to the categorical variables in our data. It helps to find out whether a difference between two categorical variables is due to chance or a relationship between them.

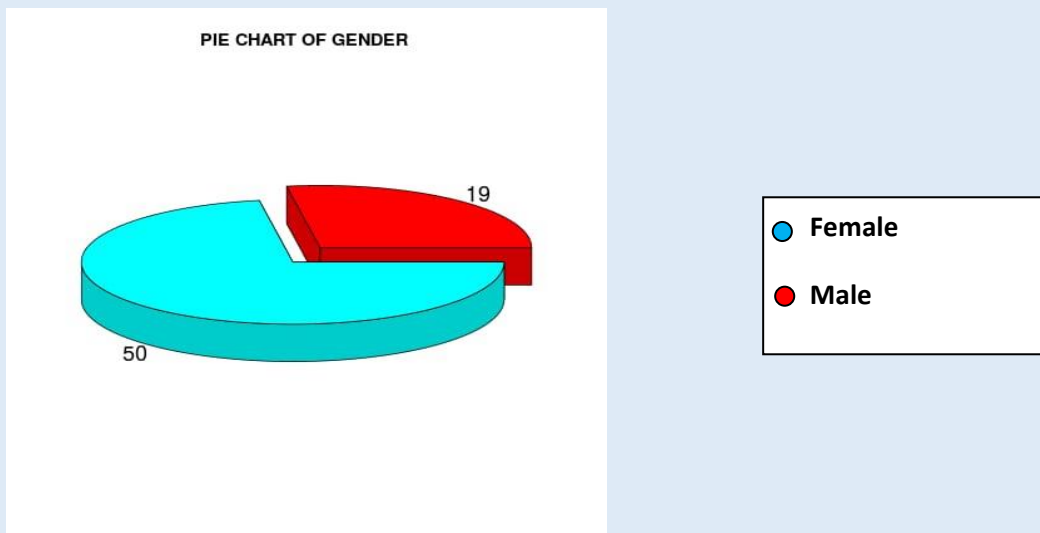
Formula

The chi-squared test is done to check if there is any difference between the observed value and expected value. The formula for chi-square can be written as;

$$\chi^2 = \sum \frac{(\text{Observed value} - \text{Expected value})^2}{\text{Expected value}}$$

4) **RESULT AND DISCUSSIONS**

- In the survey we received 69 responses from College or University students. Most of the responders were belongs to female category (72.5%). Only 27.5% of males participated in this survey.



R SCRIPT USED FOR THE PIE CHART OF GENDER

```
# Get the library.
library(plotrix)

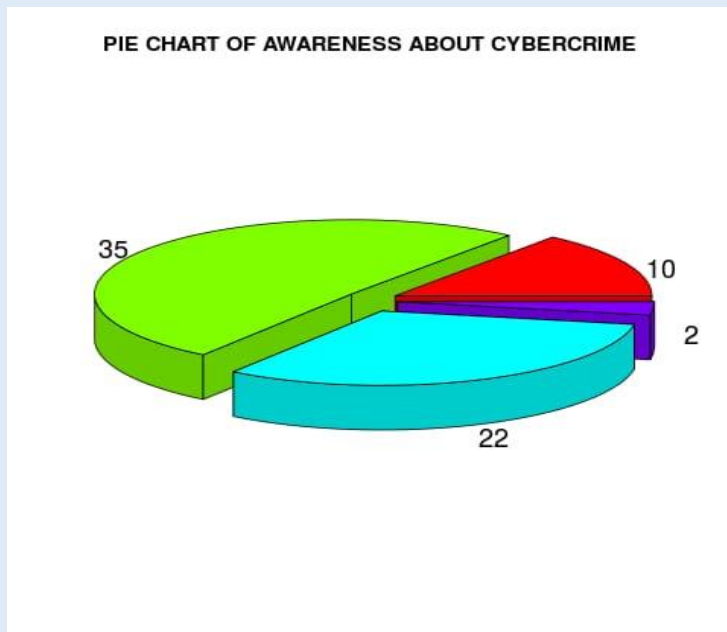
# Create data for the graph.
x <- c(19,50)
lbl <- c("Male","Female")

# Give the chart file a name.
png(file = "3d_pie_chart.jpg")

# Plot the chart.
pie3D(x,labels = lbl,explode = 0.1, main = "PIE CHART OF GENDER ")

# Save the file.
dev.off()
```

- From the survey many of the students are aware about cybercrime. Out of the 69 samples collected 10 students very well aware about it. 35 students know about it, 22 of them not so aware about it and 2 students are unaware about cybercrimes.



- Very well
- I know about it
- Not so well
- Don't know

Most of the students need more awareness about cybercrime. Only 14.5% very well about it.

R SCRIPT USED FOR THE PIE CHART

```
# Get the library.
library(plotrix)

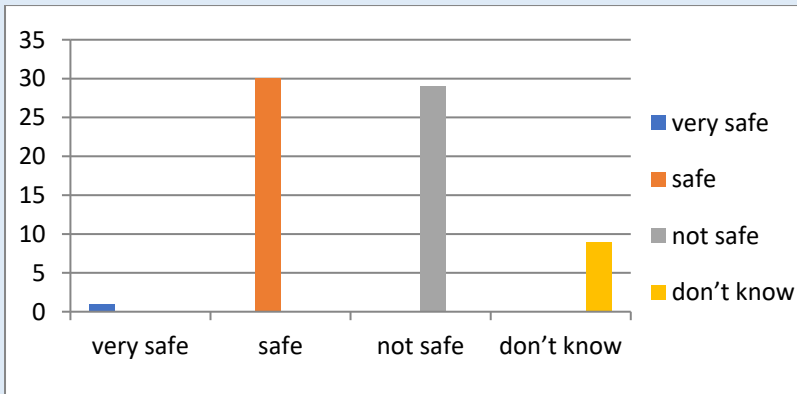
# Create data for the graph.
x <- c(10,35,22,2)
lbl <- c("very well","know about it","not so well","don't know")

# Give the chart file a name.
png(file = "3d_pie_chart.jpg")

# Plot the chart.
pie3D(x,labels = lbl,explode = 0.1, main = "PIE CHART OF AWARENESS ABOUT CYBERCRIME ")

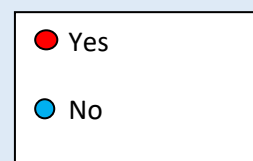
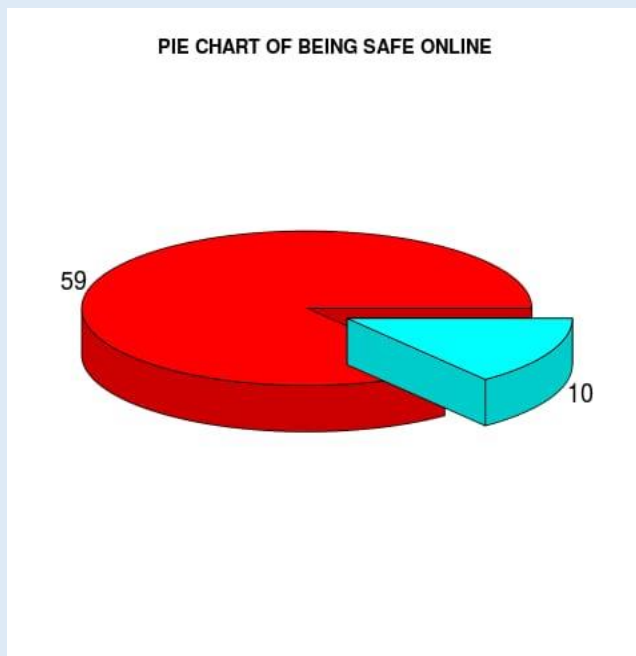
# Save the file.
dev.off()
```

- Out of the 69 samples only one student feel secure about the information when they are online. 30,29,9 responded that they feel safe, not safe, don't know about the security of their information when they are online.



50% students feel secure when they are online. 50% students don't feel any secure about their information when they are online.

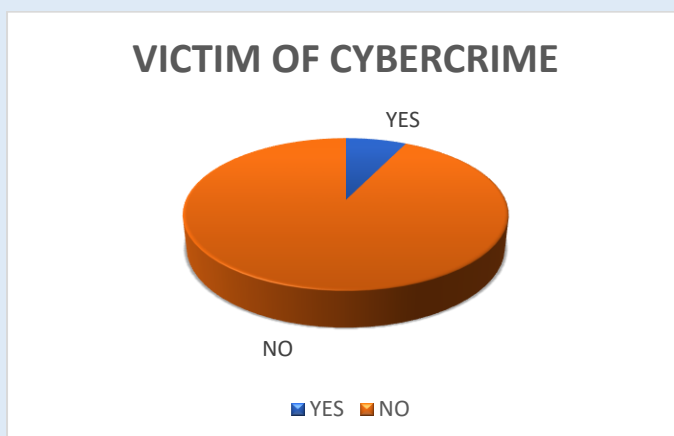
- 59 students responded that being safe online is essential. But 10 students responded it is not a required one.



R SCRIPT USED FOR THE PIE CHART

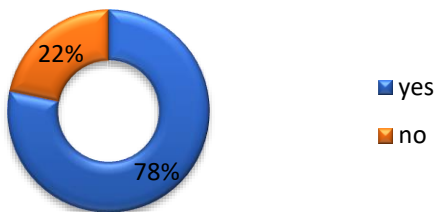
```
# Get the library.  
library(plotrix)  
  
# Create data for the graph.  
x <- c(59,10)  
lbl <- c("Yes","No")  
  
# Give the chart file a name.  
png(file = "3d_pie_chart.jpg")  
  
# Plot the chart.  
pie3D(x,labels = lbl,explode = 0.1, main = "PIE CHART OF BEING SAFE ONLINE ")  
  
# Save the file.  
dev.off()
```

- **Out of the 69 samples collected 64 students never been a victim of cybercrime. 5 of them are victim cybercrime.**



- From the 69 samples, parents of 54 students take effective measures to keep their children safe online. Actions taken by 15 parents are not effective.

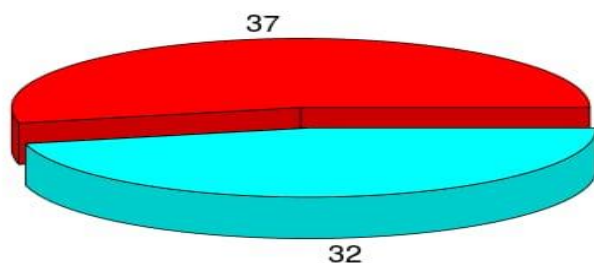
steps taken by parents to keep safe online effective



90% students get effective safety measures from their parents to keep themselves safe online.

- Out of the 69 samples 37 mentioned that in INDIA, laws are able to control cybercrime and 32 mentioned that laws are not able.

PIE CHART OF LAWS ARE ABLE TO CONTROL CYBERCRIME



● Yes
● No

50% students says laws are able to control cybercrimes

R SCRIPT USED FOR THE PIE CHART

Get the library.

```
library(plotrix)
```

Create data for the graph.

```
x <- c(37,32)
```

```
lbl <- c("YES","NO")
```

Give the chart file a name.

```
png(file = "3d_pie_chart.jpg")
```

Plot the chart.

```
pie3D(x,labels = lbl,explode = 0.1, main = "PIE CHART OF LAES ARE ABLE TO CONTROL CYBERCRIME ")
```

Save the file.

```
dev.off()
```

5) DATA ANALYSIS

HERE THE CHI SQUARE TEST FOR INDEPENDENCE IS CONDUCTED TO DETERMINE WHETHER THERE IS ANY ASSOCIATION BETWEEN THE PARAMETERS LIKE “GENDER AND AWARENESS ABOUT CYBERCRIME” ; “GENDER AND INVOLVEMENT IN CYBERCRIME” ; “AWARENESS AND VICTIM OF CYBERCRIME”.

1. “GENDER” AND “AWARENESS ABOUT CYBERCRIME”

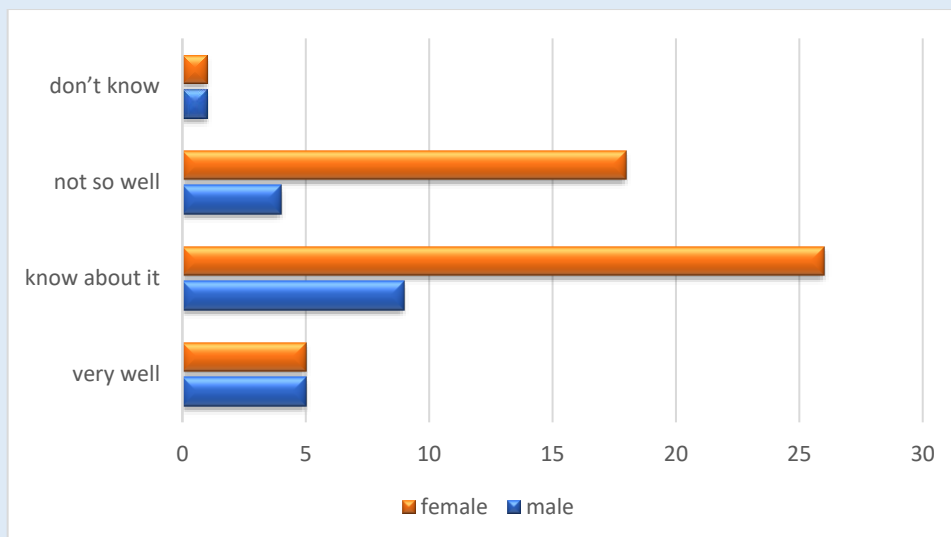
HERE OUR NULL AND ALTERNATE HYPOTHESIS ARE

H₀: PARAMETERS ARE INDEPENDENT

H₁: PARAMETERS ARE DEPENDENT

The observed frequencies of the parameters “**GENDER**” AND “**AWARENESS ABOUT CYBERCRIME**” are given below:

Awareness about cybercrime	Very well	I know about	Not so well	Don't know	Total
Gender					
Male	5	9	4	1	19
Female	5	26	18	1	50
Total	10	35	22	2	69



The expected frequencies are calculated using the equation,

$$E_{ij} = \frac{(i^{\text{th}} \text{ row total}) (j^{\text{th}} \text{ column total})}{\text{Grand total}}$$

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

No of rows (R) = 2

No of columns(K) = 4

Therefore, degrees of freedom = 3

R SCRIPT FOR X-SQUARE TEST

```
> awareness <- matrix(c(5,9,4,1,5,26,18,1), byrow = T, nrow = 2)
> awareness
      [,1] [,2] [,3] [,4]
[1,]    5    9    4    1
[2,]    5   26   18    1
> rownames(awareness) <- c("Male", "Female")
> colnames(awareness) <- c("Very well", "I know", "Not so well", "Dont know")
```

```

>
> awareness
      Very well I know Not so well Dont know
Male           5         9           4         1
Female          5        26          18         1
> model <- chisq.test(awareness)
Warning message:
In chisq.test(awareness) : Chi-squared approximation may be incorrect
> model

```

Pearson's Chi-squared test

```

data: awareness
X-squared = 4.0577, df = 3, p-value = 0.2553

```

USING R SOFTWARE WE GET THE VALUE OF X^2 WITH DEGREES OF FREEDOM 3,

$X^2(3) = 4.0577$

FROM THE X^2 TABLES FOR $\alpha = 0.05$, $X^2_{\alpha} = 7.815$

$X^2(3) < X^2_{\alpha}$ SINCE THE CALCULATED VALUE IS LESS THAN THE TABLE VALUE WE ACCEPT THE NULL HYPOTHESIS ie, WE REJECT THE ALTERNATE HYPOTHESIS.

Hence the awareness about
cybercrime is independent of
gender.

2. **“GENDER” AND “INVOLVEMENT IN CYBERCRIME”**

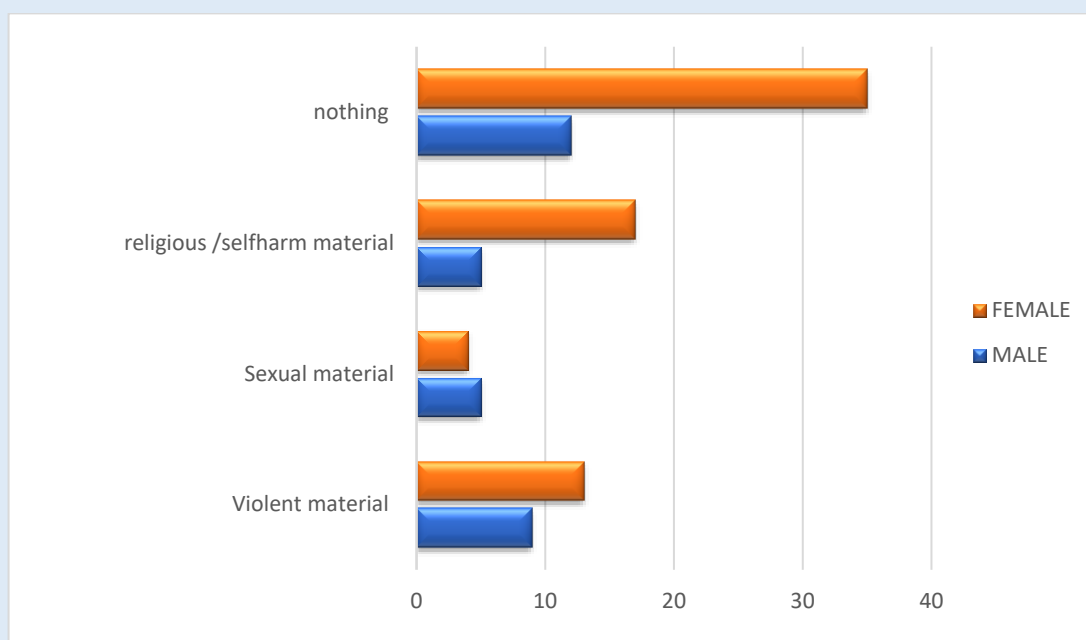
HERE OUR NULL AND ALTERNATE HYPOTHESIS ARE,

H0: PARAMETERS ARE INDEPENDENT

H1: PARAMETERS ARE DEPENDENT

The observed frequencies of the parameters **“GENDER” AND “INVOLVEMENT IN CYBERCRIME”** are given below:

Involvement in cybercrime	Violent material	Sexual material	Religious /self harm material	Nothing	Total
Gender					
Male	9	5	5	12	31
Female	13	4	17	35	69
total	22	9	22	47	100



The expected frequencies are calculated using the equation,

$$E_{ij} = \frac{(\text{i}^{\text{th}} \text{ row total}) (\text{j}^{\text{th}} \text{ column total})}{\text{Grand total}}$$

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

No of rows (R) = 2

No of columns(K) = 4

Therefore, degrees of freedom = 3

R SCRIPT FOR X-SQUARE TEST

```
> involvement <- matrix(c(9,5,5,12,13,4,17,35), byrow = T, nrow = 2)
> invovement
Error: object 'invovement' not found
> involvement <- matrix(c(9,5,5,12,13,4,17,35), byrow = T, nrow = 2)
> involvement
      [,1] [,2] [,3] [,4]
[1,]    9    5    5   12
[2,]   13    4   17   35
> rownames(involvement) <- c("Male", "Female")
> colnames(involvement) <- c("violentmaterial", "sexualmaterial",
"religiousmaterial", "nothing")
>
> involvement
      violentmaterial sexualmaterial religiousmaterial nothing
Male                9                5                5       12
Female             13                4                17       35
> model <- chisq.test(involvement)
Warning message:
In chisq.test(involvement) : Chi-squared approximation may be incorrect
> model
```

Pearson's Chi-squared test

data: involvement

X-squared = 4.9079, df = 3, p-value = 0.1787

USING R SOFTWARE WE GET THE VALUE OF X^2 WITH DEGREES OF FREEDOM 3,

$$X^2(3) = 4.9079$$

FROM THE X^2 TABLES FOR $\alpha = 0.05$, $X^2_{\alpha} = 7.815$

$X^2(3) < X^2_{\alpha}$ SINCE THE CALCULATED VALUE IS LESS THAN THE TABLE VALUE WE ACCEPT THE NULL HYPOTHESIS i.e, WE REJECT THE ALTERNATE HYPOTHESIS.

Hence the involvement in
cybercrime is independent of
gender

3. **“AWARENESS” AND “VICTIM OF A CYBERCRIME”**

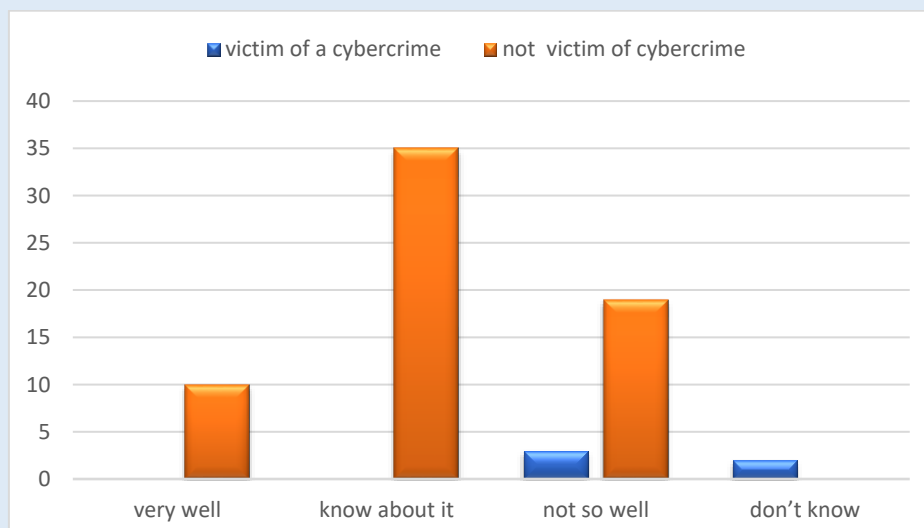
HERE OUR NULL AND ALTERNATE HYPOTHESIS ARE,

H0: THE PARAMETERS ARE INDEPENDENT

H1: THE PARAMETERS ARE DEPENDENT

The observed frequencies of the parameters **“AWARENESS” AND “VICTIM OF A CYBERCRIME”** are given below:

Victim of a cybercrime	Yes	No	Total
Awareness about cybercrime			
Very well	0	10	10
Know about it	0	35	35
Not so well	3	19	22
Don't know	2	0	2
Total	5	64	69



The expected frequencies are calculated using the equation,

$$E_{ij} = \frac{(\text{i}^{\text{th}} \text{ row total}) (\text{j}^{\text{th}} \text{ column total})}{\text{Grand total}}$$

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

No of rows (R) = 4

No of columns(K) = 2

Therefore, degrees of freedom = 3

R SCRIPT FOR X-SQUARE TEST

```
> students <- matrix(c(0,10,0,35,3,19,2,0), byrow = T, nrow = 4)
> students
      [,1] [,2]
[1,]   0  10
[2,]   0  35
[3,]   3  19
[4,]   2   0
> rownames(students) <- c("very well", "know about it", "not so well", "dont know")
> colnames(students) <- c("yes", "no")
> students
      Yes no
very well    0 10
know about it 0 35
not so well   3 19
dont know    2  0
> model <- chisq.test(students)
Warning message:
In chisq.test(students) : Chi-squared approximation may be incorrect
> model
```

Pearson's Chi-squared test

data: students

X-squared = 30.452, df = 3, p-value = 1.109e-06

USING R SOFTWARE WE GET THE VALUE OF X^2 WITH DEGREES OF FREEDOM 3,

$$X^2(3) = 30.452$$

FROM THE X^2 TABLES FOR $\alpha = 0.05$, $X^2_{\alpha} = 7.815$

$X^2(3) > X^2_{\alpha}$ SINCE THE CALCULATED VALUE IS GREATER THAN THE TABLE VALUE WE REJECT THE NULL HYPOTHESIS i.e, WE ACCEPT THE ALTERNATE HYPOTHESIS.

Being a victim of cybercrime is dependent on the awareness about cybercrime.

6) CONCLUSION

Norton survey reveals 59% Indians have dealt with cybercrime in past 12 months. A new report has stated that over 27 million Indian adults (59 per cent) have experienced identity theft in the past year. According to this we conducted a survey to know how youth aware about the cybercrime, and to circulate information and raise awareness about cyber security and cybercrime.

The study allows us to understand the importance of awareness about cybercrime among youngsters. The purpose of this is to study the trends in cybercrime and the steps that can be taken to address the rapidly growing problem.

From the survey 90 percentage of the students haven't a better awareness about cybercrime but most of them take safety measures to keep safe themselves online. The measures taken by the parents are also effective. This study actively demonstrates gender is independent of cybercrime awareness and involvement in cybercrime. But victim of a cybercrime is dependent of cybercrime awareness. So summing up we need a better awareness to keep safe ourselves from cybercrime.

It can be concluded that cybercrime is much more dreadful and disastrous in comparison to the conventional crimes. In the current scenario of development which is strictly dependent on computers, every nation and the national will have to be equipped with the knowledge of cybercrime, criminal psychology and the rules associated with it. "The best way to save you from a hacker is to hack yourself".

7) **REFERENCES**

- https://www.researchgate.net/publication/259761802_A_Survey_of_Cyber_Crime_in_India
- <https://www.thehindu.com/sci-tech/technology/52-indian-adults-dont-know-how-to-protect-themselves-from-cybercrime-survey/article34357855.ece>
- <https://medcraveonline.com/FRCIJ/cybercrime-youth-awareness-survey-in-delhi-ncr-india.html>
- <https://blog.ipleaders.in/critical-analysis-cybercrime-india/>
- <https://www.legalserviceindia.com/legal/article-4998-cyber-crime-in-india-an-overview.html>
- www.wikipedia.org
- <https://www.ijsrd.com/papers/ijsrd54.pdf>
- <https://medcraveonline.com/FRCIJ/cybercrime-youth-awareness-survey-in-delhi-ncr-india.html>
- <https://www.legalserviceindia.com/legal/article-4901-cyber-crime-among-the-youths.html>
- <https://www.skillsyouneed.com/num/graphs-charts.html>
- <https://www.investopedia.com/terms/c/chi-square-statistic.asp>
- <https://byjus.com/maths/types-of-graphs/>
- <http://ignited.in/l/a/304288>
- https://www.researchgate.net/publication/344781672_Cybercrime_Youth_Awareness_Survey_in_Delhi_NC_R_India